

# Personlighetstestning vid antagning till FHS officersutbildning

*Lennart Sjöberg*  
*Charlotte Bäckman*  
*Björn Gustavsson*

FÖRSVARSHÖGSKOLAN  
Institutionen för ledarskap och management

Försvvarshögskolan, Institutionen för ledarskap och management  
ILM Serie T:39, 2011

Serie T: Utredningar och tekniska rapporter, ISSN 1401-565X

Tryckt av One Digitaltryck i Karlstad AB, 2011

© Författarna och Försvvarshögskolans Institution för ledarskap och management

# Innehållsförteckning

<b>SAMMANFATTNING .....</b>	<b>1</b>
<b>INLEDNING .....</b>	<b>1</b>
ALLMÄNT OM PERSONLIGHETSTESTER OCH ARBETSRESULTAT .....	2
PRINCIPEN OM PROXYVALIDERING .....	3
METODASPEKTER VID URVAL OCH SJÄLVSKATTNINGSFÖRMULÄR .....	4
CTI - COMMANDER TRAIT INVENTORY .....	5
<i>CTI:s validitet och reliabilitet</i> .....	7
UPP-TESTET .....	8
<i>Narcissism</i> .....	12
<i>Validering av UPP</i> .....	12
<i>Skäl att pröva UPP-testet</i> .....	13
<b>METOD .....</b>	<b>13</b>
<b>RESULTAT .....</b>	<b>14</b>
<i>Stämningläge vid UPP-testningen</i> .....	14
<i>Skön målning</i> .....	14
<i>Datakvalitet</i> .....	18
DEN TESTADE GRUPPENS LÄGE I UPP-VARIABLERNÄ .....	19
<i>Attityd till testet</i> .....	20
<i>Reliabilitet</i> .....	21
<i>Validitet</i> .....	22
<i>Analys av narcissism</i> .....	24
<i>Personlighet och begåvning</i> .....	25
<b>REFERENSER.....</b>	<b>28</b>
<b>APPENDIX 1 .....</b>	<b>33</b>
<b>APPENDIX 2 .....</b>	<b>40</b>



## Sammanfattning

Två personlighetstester, CTI och *UPP*, gavs till en grupp sökande till officersutbildning vid FHS, sammanlagt 216 personer. *UPP*-testet är ett nytt test som presenteras i detalj i rapporten. Det har flera unika egenskaper, bl. a korrektion för skönmålning, mätning av datakvalitet och mätning av emotionellt tillstånd vid testningen. *UPP* mäter ett antal "smala" personlighetsvariabler av särskilt intresse för arbetslivet - förutom de traditionella dimensionerna i femfaktormodellen - och har i tidigare studier visat sig ha god validitet mot arbetsresultat. *UPP* mäter även arbetsmotivation och relaterade variabler som används som kriterier i denna studie. *UPP* har även genomgått STP-prövning. CTI är ett äldre test utvecklat för officersprövningar. Det bygger till stor del på Jungs typ teori. Resultaten visar att *UPP* bedömdes i genomsnitt mycket positivt av de testade personerna. Motsvarande data för CTI saknas. *UPP*-skalorna hade god reliabilitet, men de hade högre korrelation med skönmålning än CTI-skalorna. Efter korrektion för skönmålning korrelerades testen med arbetsmotivation och relaterade dimensioner. *UPP* förklarade ca tre gånger större andel av variansen hos dessa än vad CTI gjorde. Ett index som byggde på de mest kraftfulla skalorna i *UPP* hade en validitet mot de sammanslagna proxykriterierna på 0,65 efter korrektion för mätfel i kriteriet.

## Inledning

Försvårshögskolan har på uppdrag av Försvårsmakten att kontinuerligt följa upp och utvärdera framförallt de test som används vid officersurvalet. Officersurvalet utförs av Specialurval (Rekryteringsavdelningen, Specialistenheten, HRC) (tidigare Uttagningskommissionen, UTK, HRC/ARA) som 1999 blev en försvårsmaktsgemensam resurs för urval till grundläggande officersutbildning. I gemensam diskussion med urvalspsykologerna vid UTK togs 2009 ett beslut om att undersöka alternativa personlighetstest till officersurvalet. Syftet var att se om det hade framkommit nya personlighetstest och resultat som skulle kunna användas för att ytterligare förbättra urvalet till officersutbildningen.

Officersurvalet som helhet bygger på en sammanvägd lämplighetsbedömning mot officersyrket, det vill säga en prognos mot lämplighet för officersyrket - med allt vad det yrket innebär av krav och påfrestningar, nationellt och internationellt. Lämplighetsbedömningen är en helhetsbedömning av individen som görs av en uttagningskommission där både psykologer och officerare ingår och grundar sig på både tester av personliga egenskaper, såsom personlighet och begåvning, och personliga samtal om motiv och motivation. Det personlighetstest som används idag, *Commander Trait Inventory* (CTI), utvecklades under 1990-talet (Carlstedt & Widén, 1998). CTI är framtaget specifikt för officersurvalet och har senare validerats avseende uppvisade beteenden (Carlstedt, 2002). En närmare beskrivning av CTI och dess relationer med uppvisade beteenden ges nedan; för djupare kunskaper hänvisas till Försvårshögskolans rapporter (Carlstedt, 2002; Carlstedt & Widén, 1998).

Ett av de nya personlighetstest som finns på marknaden idag är *Understanding Personal Potential* (Sjöberg, 2010b) och till skillnad från de flesta andra personlighetstester är det framtaget och validerat i Sverige. Det har med andra ord, liksom CTI, därigenom en fördel mot andra etablerade personlighetstester som finns på marknaden. Ett problem med importerade test som ofta förbises är att tester framtagna i exempelvis USA kan medföra språkliga och begreppsliga svårigheter som kan vara förvånansvärt stora; se t ex Paunen & Ashton (1998). En annan svårighet är att många av dessa tester har skyddade varumärken vilket omöjliggör anpassningar till den aktuella populationen, liksom forskning om testresultaten. Dessutom har testerna oftast så kallad dold kodning, vilket innebär att oberoende forskning och validering av instrumentet försvåras (Goldberg, et al., 2006).

*UPP* har därutöver vissa unika fördelar som gör det speciellt intressant att pröva. Det finns för testet en validerad metodik för att korrigera för skönmålning, som ju är ett stort problem vid användning av självrapporttest i ”skarpt läge”, se Appendix 1. Vidare har testet inbyggd mätning av arbetsrelaterade variabler som arbetsvilja. Dessa har intresse i sig och kan även användas som kriterier vid testvalidering, en användning som kommer att närmare förklaras i denna rapport. Testning med *UPP* innebär också att man tar fram data på emotionellt tillstånd vid testningen, datakvalitet samt attityd till testet när det väl är genomgången. Allt detta är information av betydelse vid tolkning av testresultatet.

*UPP*-testet mäter ett antal variabler som är fokuserade på arbetslivet men också den mera allmänna så kallade femfaktormodellen (FFM) som sannolikt är den mest studerade personlighetsteorin idag (John & Srivastava, 1999). FFM bygger på de fem personlighetsdimensionerna Mental öppenhet (*Mental Openness*), Samvetsgrannhet (*Conscientiousness*; översätts ofta till Målmedvetenhet eller Noggrannhet), Utåtrikning (*Extraversion*), Följsamhet (*Agreeableness*; översätts ofta till Vänlighet), och Emotionell stabilitet (omvänd *Neuroticism*). Se t ex Bäckman & Carlstedt (2010) för en mer översiktlig beskrivning. Förutom FFM-dimensionerna tar *UPP*-testet hänsyn till flera viktiga aspekter av en urvalssituation, till exempel skönmålning, det vill säga individers benägenhet eller vilja att i olika sammanhang framställa sig själva i god dager (Sjöberg, 2009, 2011b). Det fanns därför flera anledningar till att undersöka *UPP*-testets kvaliteter i relation till CTI och den militära populationen.

### ***Allmänt om personlighetstester och arbetsresultat***

Det är en spridd uppfattning att personligheten är av stor betydelse i arbetslivet, i praktiskt taget alla befattningar (Barrick & Mount, 1991; Barrick & Mount, 2005), och inte minst för chefsjobb, inom såväl det civila (Hogan, Curphy, & Hogan, 1994) som det militära (McCormack & Mellor, 2002). Personlighetstesterna i kombination med kognitiva tester kan bidra väsentligt till urvalsprocessens prediktiva validitet (McHenry, Hough, Toquam, Hanson, & Ashworth, 1990) och detta gäller även för olika officersgrupper (Bilgiç & Sümer, 2009; Carretta, 2010; Melchers & Annen, 2010; Wang, Zou, & Zhang, 2010). Samtidigt tyder äldre resultat citerade i översiktsarbeten (Guion & Gottier, 1965; Mischel, 1968) på att sambandet mellan personlighet och arbetsresultat är svagt, endast omkring 0,3. Liknande resultat redovisas av en ofta citerad meta-analys av personliga egenskapers prognos på arbetsresultat (Morgeson, et al., 2007a) som visat att endast personlighetsdimensionen

Samvetsgrannhet tycks ha samband med arbetsresultatet ( $r = 0,3$ ). En annan välkänd studie (Barrick & Mount, 2005) visade att olika dimensioner av personligheten tycks ha olika betydelse för olika kriterier. Sökte man kriterier som var kopplade till aspekter av att ”komma överens” (*getting along*) var de bästa prediktorerna Mental stabilitet, Samvetsgrannhet och Följsamhet. Sökte man däremot kriterier som var kopplade till aspekter av att ”komma sig fram” (*getting ahead*) var de bästa prediktorerna Ambition (en aspekt/fasett av Utåtriktning) samt dimensionerna Mental stabilitet och Samvetsgrannhet. En översiktsrapport har nyligen påvisat att den prediktiva validiteten hos personlighetstester kan ligga omkring 0,5, givet att man bildar index av relevanta testvariabler och korrigerar för mätfel i kriterierna och beskuren spridning<sup>1</sup> (Sjöberg, 2010a). Betydelsen av personlighetsdimensionerna tycks alltså i mycket vara beroende av vilka kriterier som skall prediceras, hur man hanterar personlighetsskalorna och vilka aspekter på personligheten som studeras.

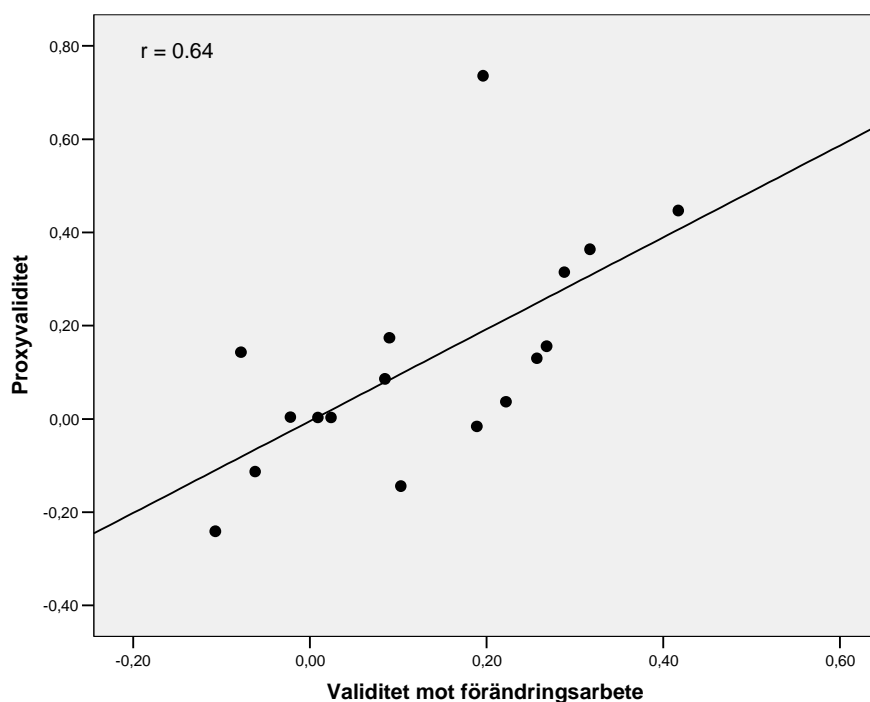
Studier har också visat att den prediktiva validiteten kan höjas om man lyckas snäva in de breda personlighetsdimensionerna till mer kriterierelevanta underskalor, se t ex Ashton, et al. (1995). Dock har arbetet med att ta fram sådana underskalor inte kommit så långt (Bäckman & Carlstedt, 2010; Dudley, Orvis, Lebiecki, & Cortina, 2006). Därtill tycks det prognostiska värdet variera beroende på vilken typ av population som undersöks (Peeters, Van Tuijl, Rutte, & Reymen, 2006). Det är alltså av mycket stor vikt att identifiera relevanta kriterier och att undersöka sambandet mellan dessa och rätt aspekt av personligheten, inom rätt population, innan man kan dra slutsatser om testinstrumentens prediktiva validitet.

### ***Principen om proxyvalidering***

Validering av tester är en dyrbar och tidskrävande aktivitet. Data på arbetsresultat måste insamlas och relateras till testresultaten. Det finns emellertid en "genväg" som kan tillgripas för att få en snabb uppfattning om ett eller flera testers förväntade validitet: proxyvalidering. Denna metod innebär validering mot kriterier som används *i stället för* (proxy) data på arbetsresultat. Proxykriterierna kan bestå av data på exempelvis arbetsvilja, resultatorientering och förändringsvilja. Sådana data samlas rutinmässigt in vid testning med *UPP*. Om man dessutom har externa kriterier kan proxyvalideringen utvärderas genom att man relaterar sambanden mot de externa kriterierna till sambanden med proxykriterierna. Figur 1 visar resultatet av en sådan undersökning. Data kommer från en studie av ledarskap där ett antal chefer inom sjukvården bedömdes med 360-graders bedömningar av sina överordnade, sidordnade och underställda (Sjöberg, Bergman, Lornudd, & Sandahl, 2011). Här ges resultat för bedömningar av chefernas funktion i förändringsarbete av deras underställda, som utgjorde det externa kriteriet medan *UPP*-testets skalor för arbetsvilja m.fl. utgjorde proxykriterier. De två serierna av validiteter har prickats i Figur 1.

---

<sup>1</sup> Det är ett vanligt problem i forskning om testvaliditet att den grupp man undersöker är selekterad med avseende på testdata. Detta beror i sin tur på att man antagit till utbildning eller anställt en delgrupp av alla sökande, och de som antas är ofta systematiskt annorlunda och mindre varierande än den ursprungliga gruppen av sökande. Om man inte korrigerar för detta förhållande kommer man att underskatta testets validitet.



**Figur 1.** Proxyvaliditeter prickade mot externa validiteter hos UPP-testets personlighets-skalar. Varje punkt i figuren motsvarar de två validiteterna hos en av skalorna.

Figuren visar en ganska god överensstämmelse mellan de två serierna av validiteter och ger stöd för användningen av proxyvalidering, inte för att ersätta validering mot externa kriterier men för att få en snabb vägledning. Andra undersökningar med UPP-testet har gett liknande resultat. I denna studie kommer principen om proxyvalidering att tillämpas på UPP och CTI.

### ***Metodasppekter vid urval och självskattningsformulär***

Det finns ett flertal andra metodasppekter på personlighetstestning som bör beaktas, men som sällan tas upp vid arbete med psykologiska tester.

**Skönmålning.** Det är fullt möjligt för en testad person att ge en överdrivet positiv (eller negativ) bild av sig själv i testet. Det är vanligt att rekryterare finner att en stor del av de testade har gett så positiva värden att testet är svårt att använda på grund av dålig differentiering. Skönmålning bör därför mätas i testet och sådana mått bör användas för att korrigera svaren så att effekten av skönmålning elimineras i största möjliga utsträckning.

**Emotionellt tillstånd vid testningen.** Om man är trött, nedstämd eller nervös kan det påverka svaren i testet och göra resultaten missvisande. Därför är det önskvärt att mäta stämningsläget vid testningen.

**Datakvalitet.** Personer som missförstår testuppgifterna eller svarar slarvigt och oengagerat ger resultat som är föga användbara. Därför är det önskvärt att mäta datakvaliteten för varje testad person.



**Attityd till testet.** De testades upplevelse av testets kvalitet och relevans är viktig för att förstå deras svar på testet, och är också en viktig aspekt på hur hela rekryteringsprocessen upplevs. Det är av stor betydelse för den rekryterande organisationen att de rutiner som används är och framstår som seriösa och vetenskapligt grundade.

Det finns således ett antal viktiga faktorer att ha kontroll på för att få fram en så ”sann” bild av den testade individen som möjligt. Lyckas man med detta kommer den prediktiva validiteten att öka avsevärt; sannolikheten för att rätt individer väljs ut kommer att öka.

### ***CTI - Commander Trait Inventory***

CTI är det personlighetstest som används vid officersurvalet till Försvarsmakten idag och utvecklades vid Försvarshögskolan (Carlstedt & Widén, 1998). CTI består av 155 påståenden som fördelar sig över 11 delskalor. Varje påstående besvaras på en sex-gradig skala från ”Instämmer inte alls” till ”Instämmer helt” och omvandlas därefter till så kallade Staninepoäng.

Delskalorna är:

1. ***Sensationsorientering***, dessa individer är utåtriktade i det att de söker handling och konkreta orsaker och sällan reflekterar över sina känslor och tankar. De har heller inget behov av att dominera andra (15 item av typen ”Jag missar sällan ett tillfälle som innebär en utmaning”; Cronbach’s alfa = 0,88)
2. ***Intuitivt beslutsfattande***, dessa individer fastnar inte i gamla beslutsmönster, utan söker hela tiden nya, upptäckta lösningar (12 item av typen ”Jag ser ofta möjligheter där andra ser svårigheter”; Cronbach’s alfa = 0,86).
3. ***Konkret tänkande***, dessa individer styrs helt av intellektuella slutsatser och objektiva data, oavsett om det rör sig om yttre fakta eller om allmänt vedertagna idéer (12 item av typen ”Konkreta fakta är det enda som gäller”; Cronbach’s alfa = 0,81).
4. ***Abstrakt tänkande***, dessa individer styrs av subjektivitet. Det är den egna känslan för saker och ting som räknas och fördjupning är målet, inte bredd (16 item av typen ”Jag ägnar en hel del tid åt att fundera och reflektera över olika saker”; Cronbach’s alfa = 0,82).
5. ***Ytlig värdeorientering***, dessa individer värderar allt utifrån erkända värderingar – inte utifrån sina egna, inre, värderingar (14 item av typen ”Jag hänger med i alla nya trender”; Cronbach’s alfa = 0,77).
6. ***Ideologisk värdeorientering***, dessa individer söker en vision och uppmärksammar inte information som strider mot, eller inte går att införliva i denna (11 item av typen ”Det är viktigt att komma fram till egna ideal och leva efter dem”; Cronbach’s alfa = 0,85).
7. ***Empati***, avser att mäta individens förmåga att leva sig in i kamraternas problem och allmänna intresse för att ta hand om dem (15 item av typen ”Jag tröstar ofta arbetskamrater som har problem”; Cronbach’s alfa = 0,90).

8. **Chefsmotivation**, avser att mäta självförtroende och viljan att ta på sig ett chefsansvar (15 item av typen ”Jag är lämplig för ledande befattningar”; Cronbach’s alfa = 0,89).
9. **Egocentricitet**, är anpassad från Levensson, Kiel och Fitzpatrick (1995) ”Primary psychopathy” (14 item av typen ”Mitt främsta mål är att skaffa så många fördelar som möjligt”; Cronbach’s alfa = 0,86).
10. **Impulsivitet**, är anpassad från Levensson et al. (1995) ”Secondary psychopathy” (14 item av typen ”Jag tappar snart intresset för saker som jag påbörjar”; Cronbach’s alfa = 0,84).
11. **Etnocentrism**, avser fördomar gentemot kvinnor och invandrare; den sistnämnda har inspirerats av Adorno, Frenkell-Brunswik, Levinson, & Sanford (1969) (9 item av typen ”Det kommer bli problem om invandrare i större utsträckning blir chefer över svenskar”; Cronbach’s alfa = 0,83).

De sex första delskalorna (1-6) är skalor som mäter kognitiv stil och har inspirerats av Jungs typologier (Jung, 1950) av extravert respektive introvert tänkande. De fem sista (7-11) delskalorna är positiva och negativa egenskaper som Carlstedt och Widén (1998) anpassat från etablerade skalor. Delskalorna Empati och Chefsmotivation ses som ”dels nödvändiga men inte tillräckliga förutsättningar för ledarskap, dels egenskaper som i sina extremer inte är acceptabla hos anställda i försvaret” (sid. 6), medan delskalorna Egocentricitet, Impulsivitet och Etnocentrism ses som oönskade egenskaper hos individen.

Dessa 11 delskalor kan användas var för sig eller aggregeras till en övergripande, generell faktor med fyra underfaktorer. Den övergripande faktorn, *Subjektiv ledarpotential*, bygger på 9 av de 11 delskalorna. Av manualen framgår att Intuitivt beslutsfattande och Chefsmotivation har högst värden i den övergripande faktorn (0,83 vardera) följt av Empati (0,59), Ideologisk värdeorientering (0,58), och Sensationsorientering (0,50). Delskalorna Intuition och Chefsmotivation återfinns endast i den generella faktorn, medan delskalorna Egocentricitet och Etnocentricitet endast går att återfinna i underfaktorerna, de laddar med andra ord inte i den övergripande faktorn (Carlstedt & Widén, 1998).

Carlstedt och Widén (1998) tar själva upp problematiken med att använda sig av en generell faktor, då en sådan ofta är förknippad med felvarians och kallar även faktorn för ”The ideal-employee factor”, eller fritt översatt ”den perfekta medarbetarfaktorn”. Exempelvis finns här effekter på variansen av deltagarnas benägenhet att svara utifrån social önskvärdhet och/eller skönmålning. Faktorn Subjektiv ledarpotential, kan alltså avspegla både en förskönande självbild och en uppriktig personbeskrivning.

De fyra underfaktorerna i modellen är *Osmidighet*, *Äventyrlighet*, *Anpasslighet* och *Opålitlighet*, och de tre första (dvs. Osmidighet, Äventyrlighet, Anpasslighet) kan enligt Carlstedt och Widén (1998) anses vara kognitiva stilar, medan den sistnämnda (dvs. Opålitlighet) kan ses som en indikation på psykopati. Alla fyra underfaktorer är att betrakta som negativa; individer med höga värden i dessa underfaktorer är att betrakta som olämpliga för officersyrket. Ur Tabell 1 framgår den modell med den övergripande faktorn och dess underfaktorer som presenteras i CTI-manualen. Delskalorna Sensationsorientering, Konkret tänkande, Empati, Impulsivitet,

Egocentricitet och Etnocentricitet återfinns i åtminstone två av underfaktorerna, även om korrelationerna mellan underfaktorerna är mycket svaga ( $r = -0,09$  och  $-0,08$ ) (Carlstedt & Widén, 1998).

**Tabell 1.** CTI testets modell av övergripande faktor och underfaktorer

	Ledar- potential	Osmidighet	Äventyrlig- het	Anpass- lighet	Opålitlig- het
<b>Delskala</b>					
Sensations- orientering	0,50		0,49		0,23
Intuition	0,83				
Konkret tänkande	0,48	0,43	-0,30	0,16	
Abstrakt tänkande	0,35	0,64			
Ytlig värde- orientering	0,31			0,89	
Ideologisk värde- orientering	0,58	0,36			
Empati	0,59			0,19	-0,27
Chefsmotivation	0,83				
Impulsivitet	-0,22		0,51		0,52
Egocentricitet	(0,07)			0,25	0,89
Etnocentricitet	(0,09)			0,20	0,52

*Not.* Tabellen är återskapad från CTI manualen (Carlstedt & Widén, 1998, sid. 11)

### *CTI:s validitet och reliabilitet*

Den modell av faktorer med en övergripande generell faktor och fyra underfaktorer som presenterades ovan uppvisar inte helt acceptabla anpassningsmått (RMSEA = 0,09). RMSEA är enkelt uttryckt ett standardiserat felvärde avseende den modell som presenteras. RMSEA bör vara så lågt som möjligt. Acceptabla värden för personlighetstest brukar ligga runt 0,06 – 0,08. Carlstedt och Widén (1998) menar att den presenterade modellen har en rimligare psykologisk innebörd än de alternativa lösningarna och att senare valideringsstudier får påvisa riktigheten av detta val.

De valideringstest som har gjorts på CTI har visat att testet har en måttlig validitet. Sambanden med senare uppvisade beteenden hade acceptabla värden ( $r = 0,30 - 0,43$ ) (Carlstedt, 2002). CTI har även viss prognostisk validitet mot senare chefsutvärderingar från högre chefskurs avseende beföring till överste (Carlstedt & Widén, 1998). Här varierade sambanden beroende på om de studerade delskalornas eller faktorernas samband med varje påstående i utvärderingen (74 st.), eller delskalornas samband med chefsutvärderingens faktorer (5 st.). På det stora hela var det låga till måttliga samband på påståendenivån. Det var relativt få påståenden som analysen byggde på: CTI-delskalorna påvisade samband med 20 påståenden i chefsutvärderingen och underfaktorerna med nio påståenden i densamma. Dock bör

här noteras att denna studie gjordes på en tidigare version av CTI; därtill var även det formulär som användes för chefsutvärderingen under utprovning varför resultatet bör ses som preliminärt.

Det mått på reliabilitet som Carlstedt och Widén (1998) presenterar för delskalorna är alfakoefficienten, eller Cronbachs alfa, som ger ett mått på hur homogena påståendena inom respektive delskala är. De alfakoefficienter som finns tyder på hög intern reliabilitet, 0,77 – 0,90. Reliabilitetsmättet för faktorlösningen anges som ”determinacy” (Muthén, 1995) och påvisar enligt Carlstedt och Widén (1998) låg till acceptabel reliabilitet (0,70 till 0,95).

### ***UPP-testet***

*UPP* är huvudsakligen ett självrapporttest, men har också inslag av begåvningsmätning (emotionell intelligens), och bygger på en empirisk snarare än teoretisk ansats. En av grunderna är FFM men utöver de fem traditionella FFM-dimensionerna har åtta smalare och mera fokuserade personlighetsskalor som samarbetsvilja, uthållighet etc. lagts till. Syftet var att nå en högre validitet än vad som tycks vara möjligt med FFM. *UPP* innehåller också sex skalor som mäter arbetsmotivation och beslätade variabler. Vidare har det för *UPP* utvecklats effektiv metodik för att mäta och korrigera för skönmålning, ett av de största problemen med självrapporttest, i synnerhet vid tillämpningar i urvalssituationer.

Syftet med att konstruera *UPP*-testet var:

- Att konstruera ett svenskt test – alltså inte bara en översättning av en utländsk förlaga - som mäter FFM och andra för arbetsprestation viktiga dimensioner, samt tillåter korrektion för skönmålning.
- Att ta fram ett test användbart för både rekrytering och utvecklingsarbete inom ramen för ca 230 testuppgifter som tar 40 -50 minuter att genomföra, där det också finns möjlighet att använda endast delar genom en modulstruktur.
- Att skapa narrativa (beskrivande) verbala rapporter förutom kvantitativa profiler.
- Att utveckla metodiken för bedömning av testdatas kvalitet och testmotivation i varje enskilt fall.
- Att göra det möjligt att ta hänsyn till den testade personens emotionella tillstånd vid testningen.

De dimensioner som testet mäter framgår av Tabell 2. De fem förstnämnda är de traditionella FFM dimensionerna och de åtta sistnämnda de smalare personlighetsdimensionerna. Emotionell intelligens mäts med två olika skalor, en som bygger på självrapporter och en som bygger på prestationer i en uppgift där man ska identifiera emotioner i ansiktsuttryck.

**Tabell 2. Personlighetstestet UPP: Översikt av testdimensioner**

<b>Dimensionens namn</b>	<b>Högt värde (&gt; 5) utmärks av goda egenskaper i de flesta eller alla av nedanstående beteenden i arbetet:</b>
<i>Utåtvändhet</i>	Social, energisk, tycker om att ta kontakter och har lätt för det
<i>Emotionell stabilitet</i>	Klarar av stress och svårigheter utan att bli nervös, har ett lugnt humör, låter sig inte provoceras
<i>Öppenhet</i>	Tycker om abstrakta idéer, är öppen för nya tankar, vill gärna lära sig nytt
<i>Följsamhet</i>	Vilja och förmåga att anpassa sig efter situationens och andra människors krav.
<i>Noggrannhet</i>	Kontrollerar noga alla viktiga detaljer
<i>Samarbetsvilja</i>	Samarbetar gärna och i positiv anda med chefer och kamrater, lyssnar till andras råd, blir sällan försenad med uppgifter
<i>Kreativitet</i>	Har lätt att få nya idéer och uppslag, även av de mest originella och oväntade slag
<i>Emotionell intelligens</i>	Förstår sig på eget och andras känsloliv, kan hantera svåra emotionella lägen hos sig själv och andra
<i>Social förmåga</i>	Kan klara av att etablera och behålla goda relationer, löser sociala problem
<i>Grad av positiv grund-attityd</i>	Är i grunden positiv, hakar inte upp sig på motgångar eller på andras sociala oförmåga, sprider en god stämning omkring sig
<i>Uthållighet</i>	Ger inte upp trots misslyckanden, kan skaka av sig motgångar och ta itu med nya uppgifter i god stämning, låter sig inte nedslås
<i>Självförtroende</i>	Har ett grundläggande förtroende till den egna förmågan, tar gärna på sig nya uppgifter och utmaningar
<i>Perfektionism</i>	Mycket hög noggrannhet och upptagenhet med detaljer
<i>Arbetsvilja</i>	Arbetar hårt och koncentrerat och i god stämning, tar korta raster, är ofta upptagen av tankar om arbetet
<i>Arbetsintresse</i>	Är intresserad av sina arbetsuppgifter, går upp i dem, kan uppleva ”flow”
<i>Arbetsstillfredsställelse</i>	Trivs bra med jobbet, söker sig inte till andra uppgifter
<i>Resultatorientering</i>	Inställd på att få resultat, ser den egna och företagets prestation som något mycket viktigt
<i>Förändringsvilja</i>	Intresserad av förändringar som syftar till förbättringar av olika slag, ser möjligheten till förändringar som en intressant utmaning och som en möjlighet till kreativ problemlösning
<i>Balans arbete – övrigt liv</i>	Upplever inte att arbetet stör möjligheterna till ett rikt övrigt liv med familj och vänner – eller att familjen ställer krav som stör möjligheten att göra ett bra jobb

Valet av dessa testskalor grundar sig på att de kan anses vara faktorer av särskild vikt i arbetslivet. Det är också faktorer som ligger nära vardagligt tänkande om personligheten och därför troligen ganska lätta att förstå, vilket skiljer dem från dimensioner som inspirerats av Jungs teori. De sex dimensioner som avser arbetsmotivation och relaterade attityder är:

Arbetsvilja  
Arbetsintresse  
Arbetsstillfredsställelse  
Resultatorientering  
Förändringsvilja  
Balans arbete-övrigt liv

För tendensen att svara på ett socialt önskvärt sätt, skönmålning, används två skalor:

Direkt mätning (overt skönmålning)  
Indirekt mätning (kovert skönmålning)

Den direkta skalan (overt skönmålning) är gjord på ett traditionellt sätt och är en skala på social önskvärdhet, dvs. den mäter om man rapporterar att man utfört socialt önskvärda men ovanliga handlingar, eller förnekar att man utfört negativa, men vanliga handlingar. Den mest kända skalan av denna typ konstruerades av Crowne och Marlowe (1960). Den indirekta, koverta skalan konstruerades för att ta fram en skala som skulle vara mindre transparent, vilket ansågs önskvärt trots att den direkta skalan troligen sällan genomskådas av de testade. Den indirekta skalan togs fram från en uppsättning testuppgifter som var särskilt högt korrelerade med den direkta skalan.

Användningen av skalor för att mäta skönmålning har kritiserats (Uziel, 2010). Det är emellertid helt klart att skönmålning påverkar rangordningen av de testade. Om ingen korrektion görs gynnas dem som bluffar i sina testsvar; effekterna kan bli mycket stora i vissa fall. När man jämför testade i skarpt läge med dem som testats i oskarpt läge får man mycket stora skillnader i skönmålning, och parallellt därmed stora skillnader i personlighetsskalor. Dessa skillnader försvinner efter korrektionen, givet att grupperna är i övrigt jämförbara. Det har också visat sig att korrektion för skönmålning med den metodik som används i *UPP* kan höja validiteten hos testskalorna (Sjöberg, et al., 2011).

*UPP*-testet inleds med 12 frågor som syftar till att mätas stämningläget i tre dimensioner:

Hedonisk ton (glad-ledsen)  
Vakenhet (pigg-trött)  
Avspändhet (lugn-spänd)

Testningen avslutas med bedömning av hur bra den testade ansåg att testet var, på åtta skalor. Exempel:

Anser du att testningen var meningsfull?  
Var testuppgifterna klart formulerade?

*UPP* använder också aggregerade variabler. Fördelen med aggregerade variabler är att man arbetar med ett mindre antal dimensioner som är nära relaterade till praktiska behov. Uppdragsgivaren kan själv definiera vilka vikterna är för "lämplighet" i en viss

tillämpning. Dessa vikter används sedan för beräkning av en sammansatt poäng. I vissa fall är det även av intresse att ta fram vikter för ”olämplighet”, som också kan läggas till grund för en aggregerad skala. Faktorerna som används är:

*Jagstyrka.* Elva av delskalorna i testet avspeglar begreppet jagstyrka. Dessa är:

- Vänlighet
- Emotionell intelligens
- Öppenhet
- Emotionell stabilitet
- Extraversion
- Kreativitet
- Uthållighet
- Samarbetsvilja
- Positiv hållning
- Självförtroende
- Social förmåga

*Stresstålighet.* Stresstålighet mäts som ett medelvärde av följande testdimensioner. De ingår med positiv vikt med undantag för perfektionism:

- Emotionell intelligens
- Uthållighet
- Samarbetsvilja
- Positiv grundattityd
- Social förmåga
- Emotionell stabilitet
- Perfektionism

*Chefspotential.* Följande variabler användes för att skapa den aggregerade variabeln chefspotential:

- Emotionell intelligens
- Öppenhet
- Arbetsvilja
- Emotionell stabilitet
- Extraversion
- Arbetsintresse
- Kreativitet
- Uthållighet
- Samarbetsvilja
- Positiv grundattityd
- Resultatorientering
- Självförtroende

- Social förmåga
- Självpuppskattning/narcissism (medelhöga värden)

*UPP*-testets datorsystem skapar också narrativa rapporter som bygger på de kvantitativa testresultaten. Dessa är till hjälp vid tolkning och skapar en "bild" av den testade, som kan generera hypoteser av värde bl a vid intervjuer. *UPP*:s datorsystem skapar också för varje testad person en intervjuguide med förslag på frågor.

### *Narcissism*

I den tillämpning som rapporteras här lades ännu en personlighetskala till, i syfte att mäta narcissism. Detta är en dimension som uttrycker mycket positiv självuppfattning och dominans. Aktuell forskning tyder på att det är en dimension<sup>2</sup> med tydliga relationer till negativt beteende (O'Boyle Jr, Forsyth, Banks, & McDaniel, 2011; Spector, 2011). Det är en variabel som är av potentiell betydelse för ledarskap; extremt höga värden kan medföra stora problem (McFarlin & Sweeney, 2010), Även mycket låga värden kan vara problematiskt för en chef. Narcissism kan också ha ett positivt inflytande (Paunonen, Lönnqvist, Verkasalo, Leikas, & Nissinen, 2006). Dimensionen mäts i *UPP* med 20 testfrågor (reliabilitet = 0,73, Cronbachs alfa). Den har begreppsvaliderats mot NPI-16 (Ames, Rose, & Anderson, 2006), en etablerad skala för mätning av narcissism. Reliabilitet hos den översatta versionen av NPI-16 var 0,64. Korrelationen mellan de två måtten på narcissism var 0,63, korrigerat för mätfel 0,80. *UPP*-testets skala för mätning av narcissism är alltså reliabel och begreppsvaliderad. Beteckningen narcissism är troligen mindre lämplig i praktiskt arbete, då den kan upplevas som negativ, sannolikt för den tankarna till en personlighetsstörning. Den kan dessutom vara svår att förstå för många och termen kommer därför att bytas mot självuppskattning.

### *Validering av UPP*

En studie av ledarskap inom vården gav höga validiteter hos *UPP* mot 360-graders-bedömningar av ledarskap, i genomsnitt omkring 0,5 (Sjöberg, et al., 2011). En undersökning av anställda inom polismyndigheter gav likartade validitetsvärden mot lönekriterier använda inom myndigheterna (de Colli, 2011); ännu ett liknande resultat rapporterades från en undersökning av anställda inom kundservice hos ett finansbolag (Sjöberg, 2010c).

Dessa studier var alla inriktade mot externa kriterier. Det finns flera undersökningar som har gett goda resultat vid validering mot proxykriterier (Sjöberg, 2011a); den senaste är en studie av sökande till polishögskolan (Sjöberg & Annell, 2011), där *UPP*-data också relaterades till bedömningar baserade på intervjuer. Den sistnämnda studien är intressant då den visade att *UPP* fungerade bättre än ett alternativt "integritetstest" som bygger på tre av dimensionerna i FFM. Eftersom FFM-dimensioner sällan visat sig ha hög validitet i arbetslivet var detta resultat väntat.

---

<sup>2</sup> Narcissism är en del av den s k mörka triaden (dark triad) vars övriga komponenter är Machiavellianism (Christie & Geis, 1970) och psykopati. De två senare mäts inte i *UPP*-testet, men vi har funnit att emotionell intelligens är tydligt negativt relaterad till Machiavellianism (Engelberg & Sjöberg, 2005), varför EI kan tjäna som en proxy för den delen av den mörka triaden.



## Skäl att pröva UPP-testet

Skäl till att pröva *UPP* ligger i dess starka sidor som redovisats ovan och som kan sammanfattas på följande sätt:

- Testet mäter ”smala” variabler som är fokuserade på arbetsresultat och har visat sig ha hög validitet mot externa data på arbetsresultat, omkring 0,5.
- Korrektion för social önskvärdhet (skönmålning) – validerad i ett flertal studier.
- Mätning av datakvaliteten i varje enskild testning.
- Mätning av stämningläge vid testningens början – en faktor som kan starkt påverka resultaten och snedvrیدا dessa.
- Mätning av attityd till testet – viktigt för trovärdigheten hos rekryterings- och bedömningsprocessen.
- Testet mäter också arbetsattityder och arbetsmotivation, dessa skalor är intressanta i sig i många sammanhang men de har också med framgång kunnat användas som ”ställföreträdande kriterier” (proxykriterier) vid testvalidering.
- Kraftfull och säker IT-plattform (inga data skickas utomlands).
- Plattformen har kapacitet för 1000-tals samtidiga testningar.
- Flexibel hållning: ständig utveckling och forskning, hög psykometrisk och vetenskaplig kompetens bakom testet.
- *UPP* har också utvecklats till en screeningversion, kallad *UPP/Screen*.
- *UPP* är granskat och bedömt av STP<sup>3</sup>; STP har bedömt och godkänt testet enligt EFPA:s<sup>4</sup> samtliga övergripande kriterier. Detta krav uppfylls ej av CTI.

## Metod

Två separata datainsamlingar genomfördes. Den första gjordes våren 2010 i samband med ansökan och urval till OP 10/13 ( $n = 56$ ). De 56 som fyllde i enkäten vid detta tillfälle tog både CTI och *UPP* vid ett och samma tillfälle, även om det bara var CTI som användes för urval. Det andra tillfället var hösten 2010 vid MHS Karlberg, 159 kadetter vid OP 09/12 fyllde i enkäten vid det tillfället. För dessa 159 rådde således andra förutsättningar, då de tog *UPP* ett år efter påbörjad utbildning och cirka 18 månader efter CTI som togs vid ansökan under våren 2009. Den skillnad som föreligger mellan de två kadettkullarna kan ha viss betydelse då det rör sig om två skilda sampeltyper (ansökande vs. Antagna).

Totalt ingick alltså 215 officersaspiranter och kadetter i studien, varav 195 män och 17 kvinnor. Uppgift om kön saknas för tre av deltagarna. Deltagarnas ålder låg i intervallet 20-36 år, median 22 år. Av deltagarna sökte eller gick 198 OP och 17 SOU (uppgift om utbildningsspår saknas för 1 deltagare).

De normdata som kom till användning omfattade 1269 testade personer, huvudsakligen ungdomar med viss högskoleutbildning (normgruppen). Detta medför en viss skevhet i förhållande till populationen men demografiska data har svaga

---

<sup>3</sup> Stiftelsen för Tillämpad Psykologi, tillhör Psykologförbundet

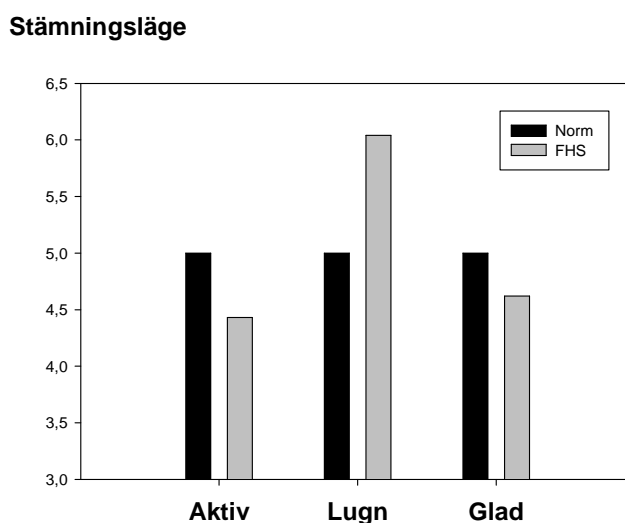
<sup>4</sup> Samarbetsorgan för de europeiska psykologförbunden

samband med testets olika skalor, och de testade vid FHS (FHS-gruppen) var i övrigt demografiskt lika normgruppen.

## Resultat

### *Stämningläge vid UPP-testningen*

Testningen med *UPP*-testet påbörjades med mätning av stämningläge. De genomsnittliga resultaten framgår av Figur 2. Dessa ges i Stanineskalor<sup>5</sup> (medelvärde,  $M = 5$ ; standardavvikelse,  $SD = 2$ ) om inget annat anges (gäller i hela resultatavsnittet).



**Figur 2.** Stämningläge vid testningens början, genomsnitt av Staninevärden.

De testade var alltså lite trötta och tydligt avspända, dessutom lite nedstämda. Samtliga skillnader mellan FHS-gruppen och normgruppen är statistiskt signifikanta.

### *Skönmålning*

Korrektionen för skönmålning görs med hjälp av regressionsmodeller, en för varje testskala. Residualerna i prognos av testskalans värde utifrån de två måtten på skönmålning är de korrigerade testvärdena, och de är okorrelerade med skönmålning. Validering av denna procedur har rapporterats på annat ställe (Sjöberg, 2009) och beskrivs i Appendix 1.

Det är nödvändigt att arbeta med olika regressionsmodeller för olika testskalor eftersom de är olika starkt korrelerade med skönmålning. Tabell 3 ger korrelationerna mellan skönmålning och *UPP*-testets skalor, Tabell 4 ger motsvarande för CTI. Korrelationen mellan de två måtten på skönmålning var 0,58.

<sup>5</sup> Strängt taget är det inte stanineskalor i egentligen mening, data har nämligen inte tvingats in i normalfördelningens form.

**Tabell 3.** Korrelationer mellan måtten på skönmålning och UPP-testets grundskalor.

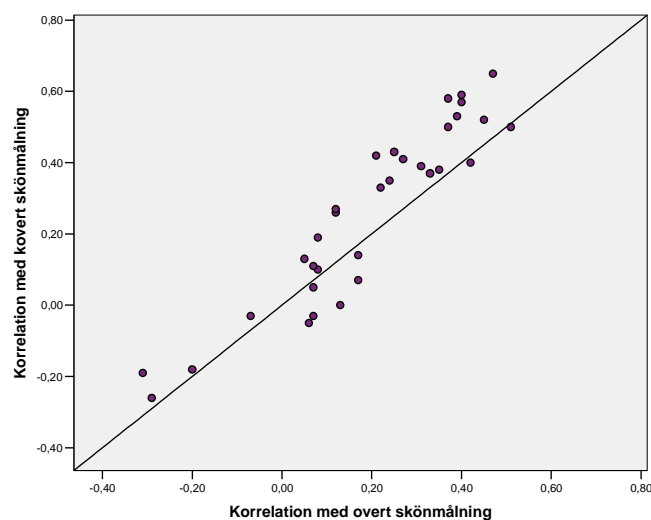
<b>Skala</b>	<b>Direkt skönmålning</b>	<b>Indirekt skönmålning</b>
Utåtvändhet	0,25	0,43
Följsamhet	0,42	0,40
Emotionell stabilitet	0,39	0,53
Öppenhet	0,22	0,33
Noggrannhet	0,51	0,50
Uthållighet	0,37	0,58
Samarbetsvilja	0,47	0,65
Positiv grundhållning	0,45	0,52
Självförtroende	0,40	0,59
Social förmåga	0,21	0,42
Emotionell intelligens, självrapport	0,40	0,57
Emotionell intelligens, emotionsbedömningar	0,07	0,11
Kreativitet	0,05	0,13
Perfektionism	0,17	0,07
Arbetsstillfredsställelse	0,12	0,26
Arbetsvilja	0,33	0,37
Resultatorientering	0,27	0,41
Förändringsvilja	0,35	0,38
Arbetsintresse	0,17	0,14
Balans arbete – övrigt liv	0,31	0,39
Hedonisk ton	0,37	0,50
Aktivitet	0,24	0,35
Spänning	0,12	0,27
Attityd till testet	0,08	0,19

**Tabell 4.** Korrelationer mellan måtten på skönmålning och CTI-testets grundskalor.

Skala	Direkt skönmålning	Indirekt skönmålning
Sensationsorientering	0,04	-0,02
Intuitivt beslutsfattande	0,21	0,13
Konkret tänkande	0,06	-0,05
Abstrakt tänkande	-0,07	-0,03
Ytlig värdeorientering	0,07	-0,03
Ideologisk värdeorientering	0,07	0,05
Empati	0,08	0,10
Chefsmotivation	0,13	0,00
Egocentricitet	-0,29	-0,26
Impulsivitet	-0,31	-0,19
Etnocentricitet	-0,20	-0,18

CTI-skalorna hade svagare korrelationer med skönmålning än skalorna i *UPP*. Detta gäller särskilt de skalor som baseras på Jungs teori. Skälet är troligen att de testade inte insåg klart vilka svar som skulle gynna dem i dessa dimensioner. De negativa sambanden med Egocentricitet, Impulsivitet och Etnocentricitet är naturliga med tanke på att det i dessa fall torde vara enklare att inse "hur man bör svara".

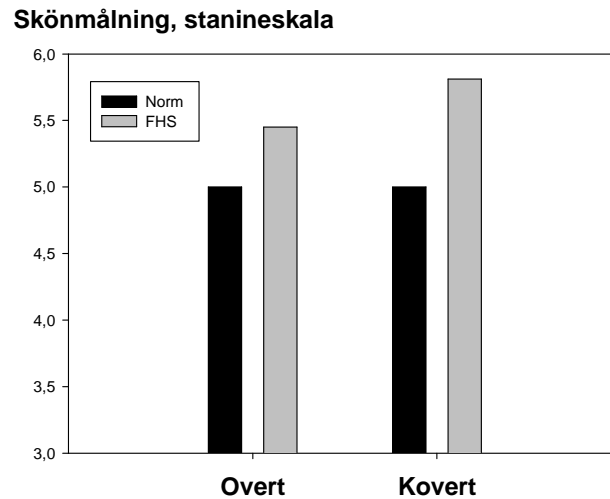
Ekvivalensen av de två måtten på skönmålning kan studeras i Figur 3, som bygger på data från båda testerna.



**Figur 3.** Korrelationer mellan testskalorna (råform) och de två måtten på skönmålning prickade mot varandra. Data från *UPP* och *CTI*.

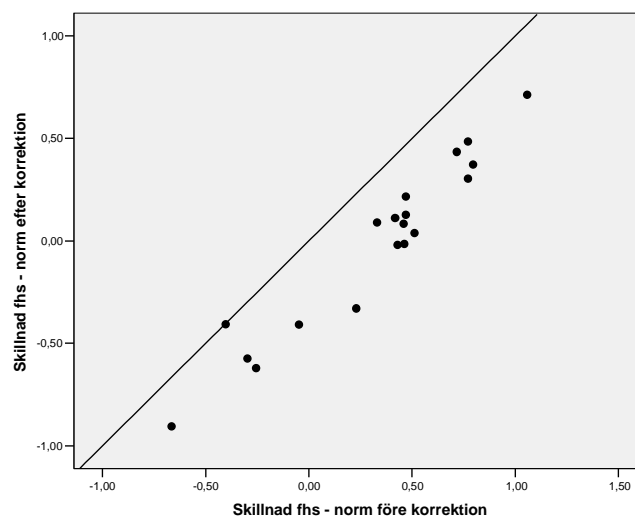
Det framgår att de två skalorna på skönmålning hade likartade samband med de övriga testskalorna, samt att dessa samband varierade kraftigt över testskalor. Det framgår också av Figur 3 att den indirekta skalan tenderade att ha starkare korrelation med övriga testskalor än vad den direkta skalan hade.

Figur 4 visar att det fanns en tydlig tendens till skönmålning i FHS-gruppen, även om den inte var extremt stark. Skillnaderna mellan normgruppen och FHS-gruppen är statistiskt signifikanta.



**Figur 4.** Direkt (overt) och indirekt (kovert) skönmålning, genomsnitt av Staninevärden.

Samtliga data korrigerades för skönmålning enligt den metodik som beskrivits ovan och i Appendix 1. Standardisering till Stanineformat gjordes sedan för rådata och korrigerade data separat. Skillnaderna mellan FHS-gruppen och genomsnitten för normgruppen illustreras i Figur 5.



**Figur 5.** Skillnad mellan normgruppen och FHS-gruppen före och efter korrektion för skönmålning.

Figur 5 visar att positiva skillnader minskat till följd av korrektionen och att negativa ökat, dvs. blivit ännu mera negativa. Den genomsnittliga skillnaden före korrektion var 1,06, dvs. en halv *SD*-enhet, vilket är en ganska betydande skillnad. Korrektionen

gjorde att den minskade till 0,71. Med andra ord försvann ca 25 % av skillnaden. Effekten av korrektion var i dessa data mindre än i tidigare studier (Sjöberg, 2009), vilket troligen beror på den måttliga nivån på skönmålningen.

## *Datakvalitet*

Datakvaliteten mäts i *UPP* med tre mått:

1. **Strukturlikhet** mäts med den intra-individuella korrelationen mellan en persons testsvar och gruppens medelvärde, beräknad över samtliga testfrågor. Denna korrelation mäter i vilken mån den testade besvarat frågorna med eftertanke och omsorg, förstått dem och uppfattat dem på ett för gruppen normalt sätt.
2. **Intra-individuell spridning** av svaren, beräknad över samtliga testfrågor. Detta spridningsmått mäter i vilken mån testpersonen besvarat frågorna nyanserat och inte mer eller mindre slentrianmässigt valt ett av svarsalternativen, t ex det mittersta.
3. **Ja-sägartendens**. Detta är ett index som beräknas som genomsnittet av svaren på positiva frågor separat och negativa frågor separat. Den som svarar jakande på positiva påståenden och nekande på negativa påståenden får ett högt värde i detta index<sup>6</sup>.

Dessa kvalitetsmått har i tidigare forskning visat sig ha god reliabilitet och stabilitet. Märk att de två förstnämnda måtten är positiva och det sistnämnda negativt när det gäller att mäta datakvalitet. I testning med *UPP* sammanvägs de till ett övergripande mått (efter vändning av måttet på Ja-sägartendens) och testrapporten innehåller en ”varning” om kvaliteten är låg.

Tabell 5 ger medelvärden (*M*) och standardavvikelser (*SD*) för kvalitetsmåten i FHS-gruppen och normgruppen.

**Tabell 5. Jämförelse i datakvalitet mellan normgruppen och data från FHS-gruppen.**

	Normgruppen		FHS-gruppen	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Strukturlikhet	0,67	0,18	0,76	0,11
Intra-individuell spridning	1,36	0,21	1,34	0,21
Ja-sägartendens	1,26	0,63	1,47	0,48

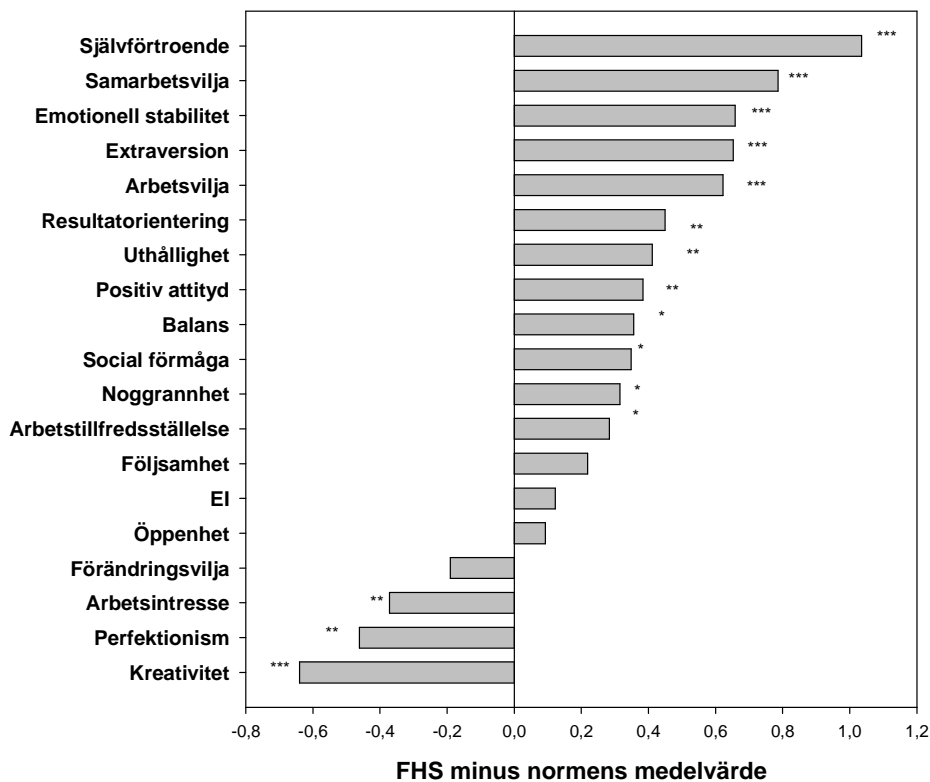
Resultaten visar jämförbar kvalitet med avseende på strukturlikhet och intra-individuell spridning<sup>7</sup>. Ja-sägartendensen var dock något högre i FHS-materialet.

<sup>6</sup> Detta mått på ja-sägartendens mäter alltså *inte* tendensen att hålla med oavsett påståendenas innehåll, på engelska *acquiescence* [5]. En sådan tendens kan i och för sig skönjas i data men har ansetts vara ointressant eftersom den är mycket svag. Det mått som används här har en mycket tydligare struktur.

<sup>7</sup> Ja-sägartendensen korrelerade i FHS-gruppen 0,53 med overt skönmålning och 0,68 med kovert skönmålning. Detta mått på Ja-sägartendens kan ses som en indirekt mätning av skönmålning och ger som korrektionsfaktor ungefär samma resultat som de två andra skalorna (overt och kovert).

## Den testade gruppens läge i UPP-variablerna

Figur 6 visar medelvärdena i UPP-testets variabler<sup>8</sup>, efter korrektion för skönmålning och standardisering till Stanineformat.



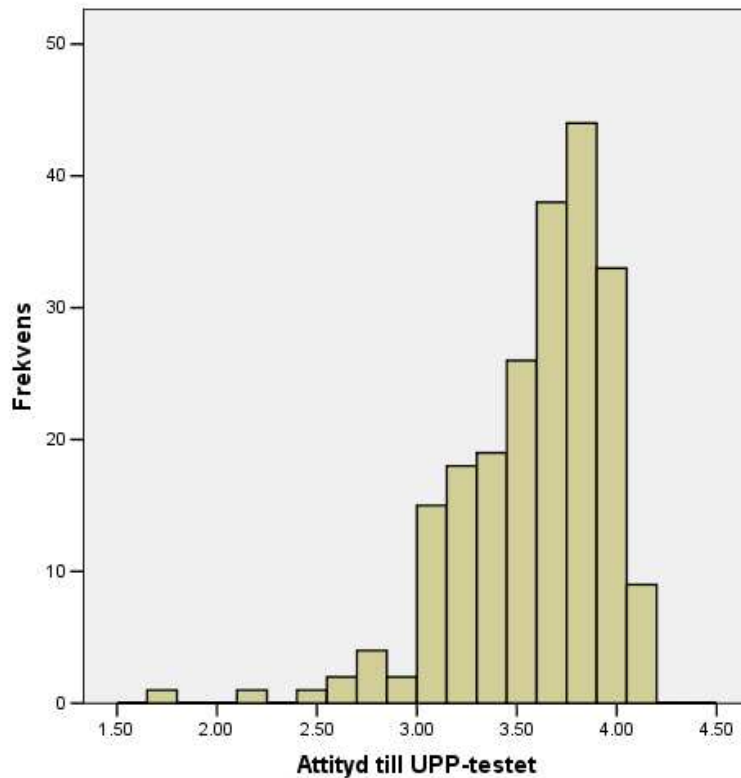
**Figur 6.** Skillnader i medelvärden för FHS-gruppen och normdata, i UPP-testets variabler, med markering för signifikansnivå i skillnaderna. Signifikansnivån anges för  $p < 0,001$  som \*\*\*, för  $p < 0,01$  med \*\* och för  $p < 0,05$  med \*.

Gruppen vid Försvarshögskolan utmärks enligt dessa data främst av Självförtroende, Samarbetsvilja och Emotionell stabilitet. De lägsta värdena finner vi i Kreativitet och därmed relaterade variabler.

<sup>8</sup> Narcissism ingår ej i denna jämförelse eftersom normdata för denna variabel saknades.

### *Attityd till testet*

Även om de testade saknar information och expertis för att bedöma ett test som de tagit är deras uppfattning viktig. Därför mäts denna attityd vid *UPP*-testning. Attityden till testet mättes med 7 frågor som slogs ihop till ett index. Fördelningen av detta index framgår av Figur 7.



**Figur 7.** *Fördelning av de testades attityder till UPP-testet.*

Figuren visar att attityderna på det hela taget, med få undantag, var klart positiva. På en särskild fråga uppgav 47 % att de ansåg att testet var bättre än andra test de tagit, lika många att det var lika bra och bara 5 % att det var sämre. Denna typ av information är sällan eller aldrig tillgänglig vid användning av andra tester, däribland CTI.



## *Reliabilitet*

Tabell 6 ger information om skalornas reliabilitet (homogenitet) i FHS-gruppen.

<b>Tabell 6. Reliabilitetsvärden för skalorna i UPP-testet, testade vid FHS (N=215)</b>			
Skala	Antal uppgifter eller ingående variabler för de aggregerade måtten	Reliabilitet (homogenitet)	Genomsnittlig korrelation mellan uppgifterna som ingår i skalan
Utåtvändhet	10	0,79	0,27
Följsamhet	8	0,63	0,18
Emotionell stabilitet	8	0,66	0,20
Öppenhet	9	0,67	0,18
Noggrannhet	9	0,72	0,23
Uthållighet	8	0,70	0,22
Samarbetsvilja	10	0,74	0,22
Positiv grundhållning	10	0,78	0,26
Självförtroende	9	0,73	0,23
Social förmåga	8	0,71	0,23
Emotionell intelligens, självrapport	18	0,78	0,18
Emotionell intelligens, emotionsbedömningar	4	0,70	0,37
Kreativitet	8	0,71	0,23
Perfektionism	9	0,70	0,20
Narcissism	20	0,73	0,10
Arbetsstillfredsställelse	3	0,79	0,56
Arbetsvilja	7	0,73	0,32
Resultatorientering	13	0,75	0,19
Förändringsvilja	8	0,73	0,24
Arbetsintresse	6	0,60	0,20
Balans arbete – övrigt liv	9	0,89	0,48
Skönmålning 1 (kovert)	12	0,65	0,14
Skönmålning 2 (overt)	9	0,70	0,21
Hedonisk ton	4	0,71	0,38
Aktivitet	4	0,73	0,41
Spänning	4	0,75	0,42
Humör, sammanslaget	12	0,64	0,37
Attityd till testet	8	0,79	0,32
Aggregerade variabler			
Jagstyrka, data korrigerade för skönmålning	11 variabler	0,70	0,18
Chefspotential, data korrigerade för skönmålning	13 variabler	0,80	0,20
Stresskänslighet, data korrigerade för skönmålning	6 variabler	0,61	0,20
Median, exklusive aggregerade variabler		0,73	0,23

## Validitet

Frågan om testernas validitet har i denna studie analyserats genom att beräkna korrelationerna med proxykriterier. Därmed avses variablerna Arbetsvilja, Arbetstillfredsställelse, Arbetsintresse, Resultatorientering, Förändringsvilja och Balans. I tidigare forskning har det dels visat sig att dessa variabler är ganska högt korrelerade med externa mått på arbetsresultat (omkring 0,4), dels att validitet mot proxykriterierna gav resultat som liknade validiteterna mot externa kriterievariabler. Denna typ av validering är därför troligen vägledande för testets värde mot funktion i arbetslivet (arbetsresultat) och har den fördelen att den kan göras snabbt, i avvaktan på mera definitiva resultat baserade på erfarenheter i arbetslivet.

I en första analys beräknades index för *UPP* och *CTI*: *UPP*-index baserades på de sex variabler som visat sig fungera bäst i tidigare studier<sup>9</sup>, samt Narcissism, alla med lika vikt. *CTI*-index baserades på de sex bästa skalorna i detta test, i betydelsen att de hade högst korrelation med proxykriterierna, utvalda på grundval av halva stickprovet<sup>10</sup>. För *CTI* ges nedan sambanden mellan testindex och kriterierna beräknade på andra halvan av stickprovet, i syfte att ge korsvaliderade resultat. Alla data har korrigerats för skönmålning som alltså inte kan förklara de samband som presenteras (se Tabell 7).

**Tabell 7. Korrelationer mellan testindex och kriterierna.**

Kriterium	<i>UPP</i>	<i>CTI</i>
Arbetsvilja	0,40	0,16
Arbetsintresse	0,25	0,27
Arbetstillfredsställelse	0,38	0,16
Resultatorientering	0,33	0,21
Förändringsvilja	0,18	0,22
Balans	0,08	-0,06
Proxykriterierna sammanslagna till ett övergripande index <sup>11</sup>	0,43 <sup>12</sup>	0,26 <sup>13</sup>

Indexet baserat på *UPP*-testet visade sig alltså ha ett starkare samband än *CTI* med det sammanslagna måttet på kriterierna. Skillnaden var stor: 28 % förklarad varians mot 10 %. Tabell 8 ger sambanden mellan det sammanslagna kriteriemåttet och aggregerade variabler i de båda testerna. *UPP*-testet arbetar med tre sådana variabler, nämligen Jagstyrka, Stresskänslighet och Chefspotential. Den sistnämnda utlämnas här eftersom den innehåller några av proxykriterierna vilket ger en falskt förhöjd nivå på sambandet med det övergripande indexet på dessa. *CTI* arbetar med fem aggregerade variabler, benämnda Osmidighet, Äventyrlighet, Anpasslighet, Opålitlighet och Ledarpotential. Se Tabell 8.

<sup>9</sup> Uthållighet, Emotionell stabilitet, Utåtvändhet, Samarbetsvilja, Positiv attityd, och Perfektionism (negativ)

<sup>10</sup> Int (intuitivt beslutsfattande), Abs (abstrakt tänkande), Ytl (ytlig värdeorientering), Ide (ideologisk värdeorientering), Emp (empati) och Lem (ledarmotivation)

<sup>11</sup> Reliabilitet 0,66

<sup>12</sup> 0,53 efter korrektion för mätfel i kriteriet

<sup>13</sup> 0,32 efter korrektion för mätfel i kriteriet

**Tabell 8.** Korrelationer mellan det sammansatta kriteriemåttet och aggregerade variabler i de två testerna.

Variabel	UPP	CTI
Jagstyrka	0,52	-
Stresskänslighet	-0,46	-
Osmidighet	-	0,05
Äventyrlighet	-	0,15
Anpasslighet	-	0,08
Opålitlighet	-	-0,01
Ledarpotential		0,22

De aggregerade CTI-variablerna tycks alltså vara svagt relaterade till kriterierna medan *UPP*-testet, särskilt på aggregerad nivå, visar på starka samband. Jagstyrkans korrelation med kriteriet når värdet 0,65 efter korrektion för mätfel i kriterieindex, mycket högt för ett personlighetstest. Man bör dock vara medveten om att proxyvaliditeter har visat sig ligga något högre än validiteter mot externa kriterier.

Sambanden mellan CTI-index enligt Tabell 7 och *UPP* var svaga:

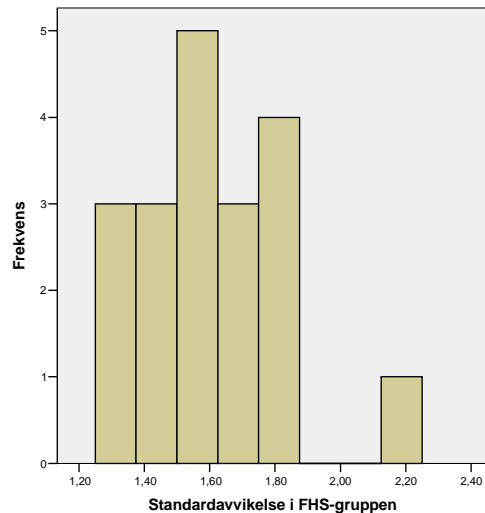
CTI-index/Jagstyrka enligt *UPP*: 0,28

CTI-index/Stresstålighet enligt *UPP*: 0,17

CTI-index/*UPP*-index: 0,10

Eftersom både CTI och *UPP* har ett aggregerad index som syftar till att mäta chefspotential, var det intressant att direkt jämföra dessa två testers respektive index för detta. Korrelationen var ganska låg ( $r = 0,34$ ;  $p < 0,05$ ) och steg bara obetydligt ( $r = 0,36$ ) om också CTI-indexet på chefspotential korrigerades för skönmålning.

Testvalidering påverkas av om spridningen i den testade gruppen är mindre än i normdata; beskuren spridning minskar den uppskattade validiteten. Figur 8 visar spridningarna i *UPP*-testets samtliga variabler.

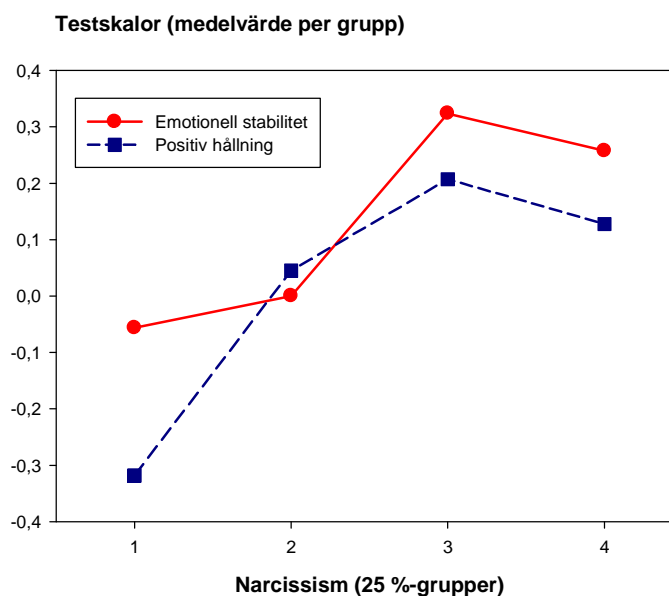


**Figur 8.** Fördelning av standardavvikelserna hos UPP-testets 21 variabler (personlighet och proxykriterier). Normgruppens värde = 2.

Det framgår att spridningarna i FHS-gruppen var betydligt mindre än i normdata, varför det är troligt att de validitetsvärden som rapporterats ovan är underskattningar.

### *Analys av narcissism*

En ny variabel för denna studie var narcissism. Den korrelerade 0,36 ( $p < 0.001$ ) (korrigerade värden) med emotionell stabilitet. Korrelationen med positiv grundinställning var också tydlig men något lägre, 0,28 ( $p < 0,05$ ). Om narcissism är en dimension med negativa effekter skulle man emellertid inte vänta sig positiva korrelationer, låt vara att de är svaga. Om man delar in materialet i fyra grupper, på grundval av nivån i narcissism, kan man dock se en väntad effekt: de som hade högsta värdet i narcissism hade lägre emotionell stabilitet och mindre positiv grundattityd än dem som hade näst högsta värdet. Se Fig. 9.



**Figur 9.** Medelvärden i två testskalor för 4 nivåer i narcissism.

Enligt Paulhus (2001) och Campbell et al. (2011) kan man förvänta att narcissism korrelerar positivt med extraversion och negativt med följsamhet. I denna studie var korrelationerna 0,35 ( $p < 0,0005$ ) och -0,16 ( $p = 0,02$ ) mellan narcissism och dessa två FFM-dimensioner.

### *Personlighet och begåvning*

Vid Försvarshögskolan har även använts ett begåvningstest, Militärhögskoleprovet (MHS-provet) och data från detta har gjort det möjligt att analysera i vilken mån *UPP* och *CTI* ger information utöver begåvning, se Tabell 9. De olika delarna i MHS-provet beskrivs i korthet i Appendix 2. Analysen av sambanden mellan MHS-provet och personlighetstesterna *UPP* och *CTI* visar genomgående på mycket svaga samband.

**Tabell 9.** Korrelationer mellan *UPP* och *CTI* och begåvningstester.

	UPP-index	Stresstålighet	Jagstyrka	CTI-index
Bongard	0,058	0,028	0,005	-0,057
Regler	0,030	0,049	0,043	-0,021
Serier	0,144	0,107	0,119	-0,094
Former	-0,020	0,015	0,008	0,053
Prickar	0,103	0,132	0,118	-0,019
Rotering	0,098	0,057	0,057	0,016
Mönster	-0,067	-0,008	-0,070	0,010
Ordkunskap	0,032	0,036	0,102	-0,048
Ordkunskap Litteratur	0,015	-0,043	0,063	0,017
Ordkunskap vetenskap	0,110	0,094	0,213	0,045
Ordkunskap teknik	0,069	0,015	0,086	-0,032
Ordkunskap debatt	0,118	0,025	0,167	0,008

Det är väl känt att begåvningstest har hög prognostisk förmåga för studie- och arbetsprestation i stort t ex (Lindqvist & Westman, 2009). Eftersom personlighetstesterna är orelaterade till dem är utsikterna goda att de ger ett betydande tillskott till prognosförmågan utöver det som uppnås med begåvningstest.

### Diskussion

*UPP*-testet hade god reliabilitet i de data som rapporterats här, och även god proxyvaliditet. *UPP*-testets validitet efter korrektion för mätfel i kriteriet var omkring 0,55 – 0,60. Fortsatt forskning behövs, inklusive långtidsuppföljning, för att dokumentera definitiv nivå på den prediktiva validiteten.

Det fanns en viss tendens till skönmålning men den försvann tack vare korrektionen. Det fanns en positiv attityd till *UPP*-testet i denna grupp. *UPP*-testet var svagt relaterat till CTI och till kognitiva tester.

Som nämndes i inledningen är det en vanlig uppfattning att personlighet är av betydelse för prestationer i arbetslivet, och inte minst torde detta gälla för militära yrken. Publicerad forskning om personlighet i detta sammanhang är inte speciellt omfattande. Amerikanska arméns "Project A" gav positiva resultat både för begåvning och för personlighet (Campbell & Knapp, 2010; McHenry, et al., 1990), men under de 20 år som gått den ursprungliga artikeln publicerades finner man bara ett fåtal publicerade militära valideringsstudier, de flesta från tillämpningar inom olika länders flygvapen. Ett undantag är en meta-analys som är orienterad mot FFM (Darr, 2011) och en studie som gett lovande resultat för smala och specifika dimensioner (Bilgiç & Sümer, 2009). De troligaste skälen till bristen på publicerade artiklar är att den forskning som utförts är belagd med sekretess, eller att resultaten inte ansetts vara av allmänt intresse ur vetenskaplig synpunkt och därför inte kunnat publiceras i tidskrifterna. Det är ju vanligt att personlighetstester används, och att de utvärderas förefaller ytterst troligt. I fortsatt forskning är det av intresse att närmare undersöka vilka aktuella erfarenheter av personlighetstestning som finns i olika länders försvarsmakter.

Bristen på publicerad forskning om militära tillämpningar står i stark kontrast med förhållandet på den civila sidan. Det har gjorts hundratals valideringsstudier av personlighetstester i arbetslivet sedan 1990 (Sjöberg, 2010a), oftast motiverade av viljan att studera variabler inom FFM. Resultaten var till en början ganska svaga, men sedan man har börjat undersöka smalare och mera fokuserade testvariabler har de blivit mera uppmuntrande. Bilden av personligheten som marginellt betydelsefull som spritts av Hunter och Schmidt (1998) är vilseledande. Ännu mera vilseledande är att vissa kända forskare velat helt underkänna personlighetstestens prognosvärde i arbetslivet (Morgeson, et al., 2007a, 2007b) – de stannade vid de övergripande FFM-dimensionerna vilka är alldeles för generella och abstrakta för att vara av värde. Det finns en allmängiltig princip som säger att psykologiska fenomen bäst förstås och prediceras av variabler som är innehållsmässigt relaterade till dem – korrelation och likhet är relaterade (Sjöberg, 1980).

Denna princip kan inte förklara de höga sambanden mellan *UPP*-testets indexvariabler och proxykriterierna; *innehållet* i variablerna har ingen överlappning, även om *metodiken* i viss mening är gemensam (skattningsskalan, instruktionerna). Det vore intressant i fortsatt forskning att undersöka i vilken mån sådana metodashänsynspåverkat korrelationerna. Variablerna i CTI, som bygger på Jungs teorier kan inte förväntas ha hög validitet eftersom de är ganska spekulativa och otillräckligt fokuserade mot arbetslivet. Omfattande forskning med Myers-Briggs Type Indicator (mest känd via sin förkortning MBTI), som är det mest kända testet grundat på Jungs typeteori (Jung, 1950), har visat att dessa dimensioner inte är valida prediktorer av arbetsresultat (Sjöberg, 2005a). Jungs dimensioner är typiska "distala" variabler, dvs. innehållsmässigt avlägsna från det man vill predicera eller förklara och det är oftast svårt att dokumentera värdet av sådana variabler; se (Sjöberg, 2003) för ett exempel.

CTI mäter också psykopati. Detta är en av tre dimensioner som har kallats "Dark Triad". De övriga två är narcissism och Machiavellianism. De två sistnämnda visade

sig i en metaanalys ha ett visst prognosvärde mot kontraproduktivt beteende i arbetssituationen (särskilt narcissism), medan psykopati inte hade det (O'Boyle et al., 2011).

Arbetsprov är i många sammanhang en bra prediktor, men ofta kan man inte få in sådana data. Ett alternativ är då simuleringsövningar, s.k. Assessment Center (AC). Dessa är mycket dyrbarare än vanliga tester: de tar lång tid (minst en dag) och kräver flera bedömare. Meta-analyser av AC-bedömningar tyder på en validitet på ungefär samma nivå som de bästa testvariablerna hos *UPP*, men inte högre, varför man kan ifrågasätta deras kostnadseffektivitet (Arthur, Day, McNelly, & Edens, 2003; Thornton Iii & Gibbons, 2009). Ett annat, traditionellt, alternativ är olika former av projektiva tester. I Sverige har flygvapnet använt ett sådant test, Defence Mechanism Test (DMT) under lång tid men tidiga positiva resultat av valideringar har inte kunnat replikeras (Sjöberg, Källmén, & Scharnberg, 1998; Martinussen, 1996; Martinussen & Torjussen, 1997). Den aktuella utvecklingen av självrapporttestning som ligger till grund för *UPP*-testet har gett mera lovande resultat.

Narcissism är en nyhet för denna undersökning. Skalan hade goda psykometriska egenskaper och god begreppsvaliditet. Som framgår av Figur 9 tycks mycket höga värden vara förknippade med ej önskade egenskaper. Variation upp till denna höga nivå gav emellertid positiva effekter. En måttlig grad av narcissism skulle därför kunna vara positiv. Campbell et al. (2001) pekar på en rad resultat där man funnit negativa samband mellan narcissism och socialt stödjande beteende i arbetsmiljön, positiva med kontraproduktiva beteenden. Sambandet med arbetsresultat tycks ligga omkring noll. Fortsatt forskning om och användning av denna personlighetsvariabel vore av stort intresse, liksom av andra negativa aspekter på ledarskap (Larsson, Fors och Nilsson, in press). Larsson et al. påpekar att negativt eller dåligt ledarskap kan vara mera negativt än vad än gott ledarskap är positivt (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). Dixons klassiska bok ger många exempel på dåligt ledarskap från höga nivåer i den militära hierarkin (Dixon, 1976).

## Referenser

- Adorno, T. W., Frenkel-Brunswick, E., Levinson, D. J., & Sanford. (1969). *The Authoritarian Personality*. New York: Norton.
- Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. [doi:10.1016/j.jrp.2005.03.002]. *Journal of Research in Personality*, 40, 440-450.
- Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, 56, 125-154.
- Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality*, 29, 432-442.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters. Moving on to more important matters. *Human Performance*, 18, 359-372.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). "Bad is stronger than good". *Review of General Psychology*, 4, 323-370.
- Bilgiç, R., & Sümer, H. C. (2009). Predicting military performance from specific personality measures: A validity study. *International Journal of Selection and Assessment*, 17, 231-238.
- Bäckman, C., & Carlstedt, B. (2010). A construct validation of a profession-focused personality questionnaire (PQ) versus the FFPI and the SIMP. [doi:10.1027/1015-5759/a000019]. *European Journal of Psychological Assessment*, 26, 136-142.
- Campbell, J. P., & Knapp, D. J. (2010). Project A: 12 years of R & D. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection*. (pp. 865-885): New York, NY, US: Routledge/Taylor & Francis Group.
- Campbell, W. K., Hoffman, B. J., Campbell, S. M., & Marchisio, G. (2011). Narcissism in organizational contexts. [doi:10.1016/j.hrmr.2010.10.007]. *Human Resource Management Review*, 21, 268-284.
- Carlstedt, L. (2002). *Beteendekorrelat till begåvnings- och personlighetsvariabler. En valideringsstudie av testen i MHS-provet och i CTI*. Karlstad: ILM, Försvarshögskolan.
- Carlstedt, L., & Widén, H. (1998). *CTI – Commander Trait Inventory. Manual*. Karlstad: LI: Försvarshögskolan.
- Carretta, T. R. (2010). Predictive validity of the Air Force Officer Qualifying Test for non-rated officer specialties. *Military Psychology*, 22, 450-464.
- Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. New York: Academic Press.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.



- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting and Clinical Psychology*, 24, 349-354.
- Darr, W. (2011). Military personality research: A meta-analysis of the Self Description Inventory. [doi:10.1080/08995605.2011.570583]. *Military Psychology*, 23, 272-296.
- de Colli, D. (2011). *Ett nytt svenskt arbetspsykologiskt test och arbetsprestation inom polisen – samtidig validitet*: Mälardalens högskola, Akademin för hållbar samhälls- och teknikutveckling.
- Dixon, N. F. (1976). *On the psychology of military incompetence*. London: Jonathan Cape.
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A Meta-Analytic Investigation of Conscientiousness in the Prediction of Job Performance: Examining the Intercorrelations and the Incremental Validity of Narrow Traits. *Journal of Applied Psychology*, 91, 40-57.
- Engelberg, E., & Sjöberg, L. (2005). Emotional intelligence and interpersonal skills. In R. D. Roberts & R. Schulze (Eds.), *International handbook of emotional intelligence* (pp. 289-308). Cambridge MA: Hogrefe.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality & Individual Differences*, 7, 385-400.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84-96.
- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18, 135-164.
- Hogan, R., Curphy, G. J., & Hogan, J. (1994). What we know about leadership: Effectiveness and personality. *American Psychologist*, 49, 493-504.
- Hough, L. M. (1998). The millennium for personality psychology: New horizons or good ole daze. *Applied Psychology: An International Review*, 47, 233-261.
- John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: history, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102-138). New York: Guilford Press.
- Jung, C. G. (1950). *Psychologische Typen*, 8. Aufl. Zürich: Rascher.
- Larsson, G., Fors, M., & Nilsson, S. (in press). Destrudo – L: development of a short scale designed to measure destructive leadership behaviours in a military context. *Leadership & Organization Development Journal*
- Lindqvist, E., & Westman, R. (2009). *The labour market returns to cognitive and non-cognitive ability: Evidence from the Swedish enlistment*. Stockholm: Research Institute of Industrial Economics.
- Lönnqvist, J.-E., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J., & Verkasalo, M. (2007). Substance and style in socially desirable responding. *Journal of Personality*, 75, 291-322.

- Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences*, *32*, 247-256.
- Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *The International Journal of Aviation Psychology*, *6*, 1-20.
- Martinussen, M., & Torjussen, T. (1997). Pilot selection in the Norwegian Air Force: A validation and meta-analysis of the test battery. [doi:10.1207/s15327108ijap0801\_2]. *International Journal of Aviation Psychology*, *8*, 33-45.
- McCormack, L., & Mellor, D. (2002). The role of personality in leadership: An application of the Five-Factor Model in the Australian military. *Military Psychology*, *14*, 179-197.
- McFarlin, D. B., & Sweeney, P. D. (2010). The corporate reflecting pool: Antecedents and consequences of narcissism in executives. In B. Schyns & T. Hansbrough (Eds.), *When leadership goes wrong: Destructive leadership, mistakes, and ethical failures*. (pp. 247-283): Greenwich, CT, US: IAP Information Age Publishing.
- McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and validity domains. *Personnel Psychology*, *43*, 335-354.
- Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, *77*, 531-552.
- Melchers, K. G., & Annen, H. (2010). Officer selection for the Swiss armed forces: An evaluation of validity and fairness issues. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie*, *69*, 105-115.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, *60*, 1029-1049.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*, 683-729.
- O'Boyle Jr, E. H., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2011). A meta-analysis of the dark triad and work behavior: A social exchange perspective. [doi:10.1037/a0025679]. *Journal of Applied Psychology; Journal of Applied Psychology* No Pagination Specified.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660-679.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social-psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.

- Paulhus, D. L. (2001). Normal narcissism: Two minimalist accounts. *Psychological Inquiry*, 12, 228-230.
- Paulhus, D. L., Bruce, N. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin*, 21, 100-108.
- Paulhus, D. L., & Reid, D. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, 60, 307-317.
- Paunen, S. V., & Ashton, M. C. (1998). The structured assessment of personality across cultures. *Journal of Cross-Cultural Psychology*, 29, 150-170.
- Paunonen, S. V., Lönnqvist, J.-E., Verkasalo, M., Leikas, S., & Nissinen, V. (2006). Narcissism and emergent leadership in military cadets. [doi:10.1016/j.leaqua.2006.06.003]. *The Leadership Quarterly*, 17, 475-486.
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research. Explanation and prediction*. New York: Holt, Rinehart and Winston.
- Peeters, M. A. G., Van Tuijl, H. F. J. M., Rutte, C. G., & Reymen, I. M. M. J. (2006). Personality and team performance: A meta-analysis. *European Journal of Personality*, 20, 377-396.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Sjöberg, L. (1980). Similarity and correlation. In E.-D. Lantermann & H. Feger (Eds.), *Similarity and choice* (pp. 70-87). Bern: Huber.
- Sjöberg, L. (2003). Distal factors in risk perception. *Journal of Risk Research*, 6, 187-211.
- Sjöberg, L. (2005a). En kritisk diskussion av Myers-Briggs testet. (A critical discussion of the Myers-Briggs test). *Organisational Theory & Practice. Scandinavian Journal of Organisational Psychology*, 15, 21-28.
- Sjöberg, L. (2005b). Självs kattningstest: Den som ljuger fräckast vinner? *Personal och Ledarskap*, 36, 22-25.
- Sjöberg, L. (2008). *Bortom Big Five: Konstruktion och validering av ett personlighetstest. (Beyond Big Five: Construction and validation of a personality test)* (SSE/EFI Working Paper Series in Business Administration No. 2008:7). Stockholm: Stockholm School of Economics.
- Sjöberg, L. (2009). *UPP-testet: Korrektion för skönmålning. Forskningsrapport 2009:3*. Stockholm: Psykologisk Metod AB.
- Sjöberg, L. (2010a). *Personlighetsdimensioners validitet i arbetslivet: teorier och empiri* (SSE/EFI Working Paper Series in Business Administration No. 2010:6). Stockholm: Stockholm School of Economics.
- Sjöberg, L. (2010b). *A third generation personality test* (SSE/EFI Working Paper Series in Business Administration No. 2010:3). Stockholm: Stockholm School of Economics.
- Sjöberg, L. (2010c). *UPP-testet och kundservice: Kriteriestudie. Forskningsrapport 2010:6*. Stockholm: Psykologisk Metod AB.
- Sjöberg, L. (2011a). *UPP-testet. Teknisk manual*. Stockholm: Psykologisk Metod AB.
- Sjöberg, L. (2011b). *Ökad testvaliditet genom korrektion för skönmålning*. Stockholm: Psykologisk Metod AB.

- Sjöberg, L., & Annell, S. (2011). *UPP/Screen vid antagning till polisutbildningen*. Stockholm: Rekryteringsmyndigheten, Rapport nr. 1.
- Sjöberg, L., Bergman, D., Lornudd, C., & Sandahl, C. (2011). *Sambandet mellan ett personlighetstest och 360-graders bedömningar av chefer i hälso- och sjukvården*. Stockholm: Karolinska Institutet, Institutionen för lärande, informatik, management och etik (LIME).
- Sjöberg, L., & Engelberg, E. (2005). Measuring and validating emotional intelligence as performance or self-report. In K. B. S. Kumar (Ed.), *Emotional intelligence. Research insights* (pp. 97-125). Hyderabad, India: ICFAI University Press.
- Sjöberg, L., Källmén, H., & Scharnberg, M. (1998). Selection for stressful jobs: Is the Defence Mechanism Test the solution? In R. Flin, M. Salas, M. Strub & L. Martin (Eds.), *Decision making under stress: Emerging themes and applications* (pp. 252-260). Aldershot: Ashgate.
- Spector, P. E. (2011). The relationship of personality to counterproductive work behavior (CWB): An integration of perspectives. *Human Resource Management Review; Human Resource Management Review*, 21, 342-352.
- Thornton III, G. C., & Gibbons, A. M. (2009). Validity of assessment centers for personnel selection. *Human Resource Management Review*, 19, 169-187.
- Uziel, L. (2010). Rethinking social desirability scales: from impression management to interpersonally oriented self-control. *Perspectives on Psychological Science*, 5, 243-262.
- Wang, F.-R., Zou, S.-H., & Zhang, Y.-L. (2010). [Validities of the Military Officer Vocational Personality Scale.]. *Chinese Journal of Clinical Psychology*, 18, 451-454.
- Vrij, A. (2000). *Detecting lies and deceit*. New York: Wiley.

# Appendix 1

Personlighetstest används i stor omfattning vid anställningar i Sverige. Det handlar nästan alltid om självrapporttest där den testade får ta ställning till ett antal påståenden och ange i vilken grad de stämmer in på honom eller henne, eller om de håller med om påståendena. Test av den typen är oftast lätta att genomskåda. Omfattande forskning har visat att många, troligen alla, självrapporttest lider av denna svaghet (Furnham, 1986; Hough, 1998). Frågan är vad resultaten på ett sådant test egentligen är värda, om svaren på testet är starkt korrelerade med skönmålning, och kraftigt påverkas av vilken betydelse testningen förväntas ha.

Forskningen visar emellertid att testen trots allt har ett visst prognosvärde för många olika typer av befattningar (Ones, Viswesvaran, & Reiss, 1996). Det låter ju bra att testens validitet finns kvar trots skönmålade svar, men det finns en hake i resonemanget. Man tar nämligen i praktiken inte alls, eller bara i mycket begränsad omfattning, *individuell* hänsyn till tendensen att svara taktiskt. På den individuella nivån kan effekterna av skönmålning bli stora.

Det dominerande sättet att hantera detta problem tycks vara att använda ipsativa (jämförande) svarsformat. Dessa leder emellertid till stora och olösta statistiska och psykometriska problem med analys av strukturen i testuppgifterna och bestämning av reliabilitet (Meade, 2004). Empirisk forskning visar dessutom att de ipsativa formaten inte effektivt motverkar tendensen till skönmålning (Martin, Bowen, & Hunt, 2002).

Två skönmålningsskalor<sup>14</sup> används i *UPP*-testet. Traditionellt mäter man skönmålning, eller tendensen att ge taktiksvår, med hjälp av en skala av den typ som Crowne och Marlowe (1960) publicerade, ofta kallad en skala för mätning av social önskvärdhet (social desirability). Uppgifterna i den skalan bygger på idén att de allra flesta någon gång betar sig på tvären mot gällande normsystem och att den som inte medger detta om sig själv antagligen svarar taktiskt. Ett exempel är: "Jag har aldrig kastat skräp på gatan." Om en person svarar taktiskt på sådana uppgifter är det rimligt att tro att även övriga svar på testet är taktiska och tillrättalagda för att ge ett positivt intryck.

Paulhus har fört vidare diskussionen om skönmålning på ett intressant sätt (Paulhus, 1991; Paulhus & Reid, 1991). Han använde samma typ av uppgifter som Crowne och Marlowe (1960) men också en ny typ som avsåg att mäta självbedrägeri. Dessa uppgifter handlar om psykologiska tillkortakommanden som då och då inträffar för alla – den som förnekar dem ljugar för sig själv. Ett exempel är: "Det är svårt för mig att stänga ute en störande tanke." Lönnqvist et al. (Lönnqvist, Paunonen, Tuulio-Henriksson, Lönnqvist, & Verkasalo, 2007) fann att Paulhus-skalorna åtminstone till en del tycktes mäta metodartefakter på det avsedda sättet. Man kan diskutera om inte samma sorts taktiksvår ligger bakom nekande svar på sådana uppgifter som på de mera traditionella uppgifterna som handlar om normbrott. Hursomhelst korrelerar de ganska högt med den traditionella typen av taktiksvår. I en urvalssituation ( $N = 282$ , sökande till Handelshögskolan) användes Crowne-Marlowes skala och Paulhus båda skalor. Paulhus skala (Paulhus, Bruce, & Trapnell, 1995; Paulhus & Reid, 1991) som byggde

---

<sup>14</sup> Termen skönmålningsskala föredras här framför lögnskala – det handlar ju om att man svarar taktiskt vilket inte behöver beskrivas med den nedsättande termen lögn. En annan vanlig beteckning är social önskvärdhet.

på frågor om normbrott korrelerade 0,84 med Crowne-Marlowe, medan skalan på ”självedrägeri” korrelerade 0,54 – också det ett högt värde.

Ett viktigt problem är frågan om uppgifterna i en taktiskskala går att genomskåda. Tag som exempel ovanstående uppgift om att kasta skräp på gatan. Det är absurt att påstå att man aldrig gör det – åtminstone för de allra flesta. Sådana uppgifter kan därför skapa misstänksamhet, och de testade personer som förberett sig genom att läsa på om testmetodik kan kanske identifiera vilka uppgifter det rör sig om och räkna ut hur de ska besvaras för att nå ett maximalt positivt resultat på testet (även om detta inte är så helt lätt att göra).

I arbetet med *UPP*-testet har konstruerats en egen skala på samma principer som Crowne-Marlowe och Paulhus, en overt skönmålningskala (öppen mätning), men också en kovert skala (dold mätning). Den senare består av uppgifter av samma typ som i den övriga delen av testet, men de är särskilt utvalda för att korrelerade högt med den overta taktiskskalan. På så sätt får man en möjlighet att mäta taktiksvår som inte kan avslöjas av sofistikerade testtagare. De två taktiskskalorna (overt och kovert) har begreppsvaliderats, se Tabell A1. Båda korrelerade högt med Crowne-Marloweskalen och högt sinsemellan. Den senare korrelationen, mellan de overta och koverta skalorna, var något lägre vilket är rimligt.

**Tabell A1.** *Korrelationer mellan de tre skönmålningsskalorna (N=159), data från Studie 1 i den tekniska manualen för UPP.*

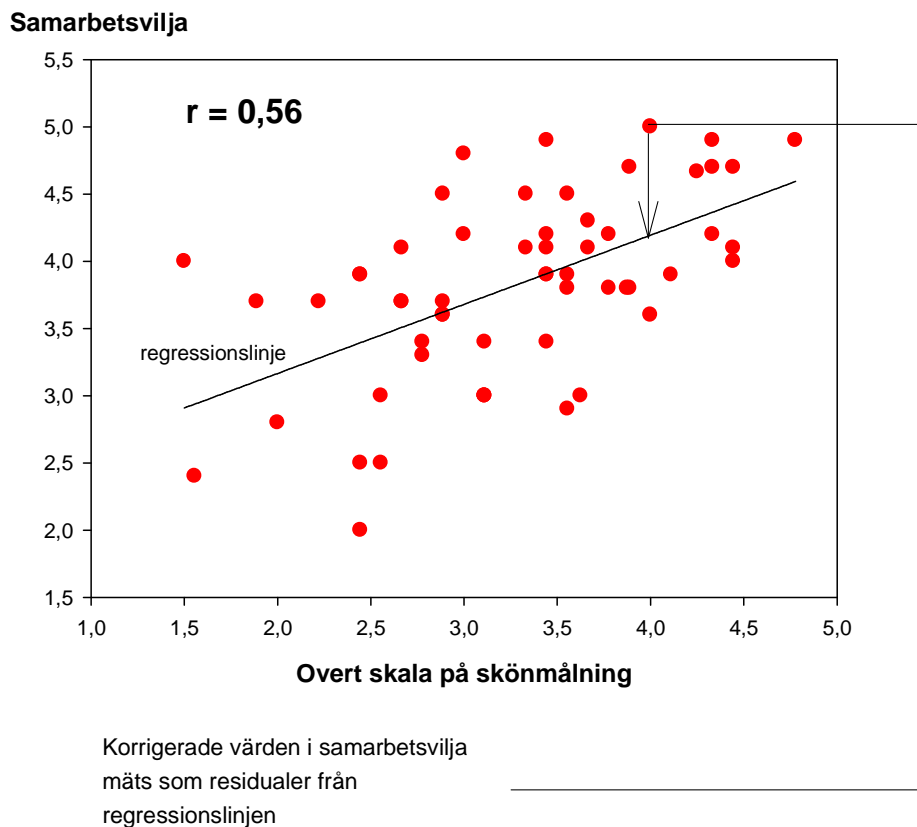
	Crowne-Marlowes skala	Overt skönmålningskala	Kovert skönmålningskala
Crowne-Marlowes skala	1,00	0,76	0,73
Overt skönmålningskala	0,76	1,00	0,56
Kovert skönmålningskala	0,73	0,56	1,00

Hur korrektionen för skönmålning görs är av stor betydelse. Det finns tre vanliga metoder, alla otillfredsställande:

1. Ipsativt format används, ibland med försök att matcha valalternativ i social önskvärdhet.
2. Psykologen som tolkar testresultatet gör en intuitiv bedömning av resultatets trovärdighet på grundval av värdet på en ”lögnskala” och sin egen bedömning av den testades ”ärlighet”.
3. Man använder sig av ett gränsvärde. De 25 % av de testade som har högst värde på taktiskskalan sällas bort och kommer inte ifråga för antagning eller anställning.

Problemen med dessa metoder är dels att forskning har visat att de inte har den avsedda effekten (Metod 1), att de är alltför subjektiva (Metod 2), eller ineffektiva (Metod 3). Att avslöja lögnare utifrån subjektiva intryck är något de flesta tror att de kan göra – men de kan det sällan (Vrij, 2000). Metoden med gränsvärde rensar bara bort en mindre del av variansen som beror på skönmålning. De som rankas högst på testet tenderar att vara bluffmakare även efter det att den metoden har använts.

Alla metoderna har den gemensamma nackdelen att de inte korrigerar testvariablerna *specifikt* utan de försöker göra en global korrektion. Den metodik som använts i *UPP*-testet innebär att korrektionen görs för varje testvariabel separat. Genomslaget för taktiksvarstendens kan ju vara mycket varierande beroende på vilken dimension det är som är aktuell. Den metodik som använts här utnyttjar effekten av taktikskalorna fullt ut med hjälp av en regressionsmodell och beräkning av residualer, se Figur A1.



**Figur A1.** Illustration av hur korrigerade värden i ett test uppskattas med hjälp av en psykometrisk modell.

Beräkningen av residualer görs enligt vanlig regressionsanalys, dvs

Residual = råvärde – predicerat värde på grundval av modellen

Formlerna blir enklast om råvärdena först har standardiserats till medelvärde = 0 och standardavvikelse = 1. För fallet med endast en oberoende variabel är regressionskoefficienten = korrelationen mellan beroende och oberoende variabel; vid flera oberoende variabler gäller samma principer men formlerna är mera komplicerade. All denna information finns i vanliga läroböcker i statistik; se t ex Cohen, Cohen, West, & Aiken (2003) för en utförlig och avancerad diskussion. Explicita formler ges sällan för fallet med två prediktorvariabler, men de kan återfinnas t ex hos Pedhazur (1982). I vårt fall tillämpas fallet med två oberoende variabler (de två skönmålningsskalorna) på följande sätt:

Modellen för en testvariabel  $y$  och de två måtten på skönmålning, som indexeras  $1$  och  $2$ , är följande:

$$z_y = \beta_1 z_1 + \beta_2 z_2 + e$$

där:

$\beta_1$  och  $\beta_2$  är de standardiserade regressionsvikterna;

$z$ -värdena indexerade  $y$ ,  $1$  och  $2$  är  $z$ -transformerade variabler av respektive testvariabeln  $y$  och de två måtten på skönmålning  $1$  och  $2$ ;

$e$  är en felterm.

Det från modellen predicerade testvärdet erhålls genom:

$$z_{ypred.12} = \beta_1 z_1 + \beta_2 z_2$$

Residualen är:

$$res_{y.12} = z_y - z_{ypred.12}$$

Residualen är testvärdet korrigerat för skönmålning, med samtidig korrektion för overt och kovert skönmålning. I fortsatta beräkningar standardiseras den i normgruppen till medelvärde = 0 och standardavvikelse = 1.

### **Estimering av regressionsvikter**

$\beta_1$  och  $\beta_2$  är de standardiserade regressionsvikterna som estimeras enligt följande formler:

$$\beta_1 = (r_{y1} - r_{y2} r_{12}) / (1 - r_{12})$$

$$\beta_2 = (r_{y2} - r_{y1} r_{12}) / (1 - r_{12})$$

Där:

$r_{y1}$  är korrelationen mellan (okorrigerade) testvariabeln  $y$  och overt skönmålning;

$r_{y2}$  är korrelationen mellan (okorrigerade) testvariabeln  $y$  och kovert skönmålning;

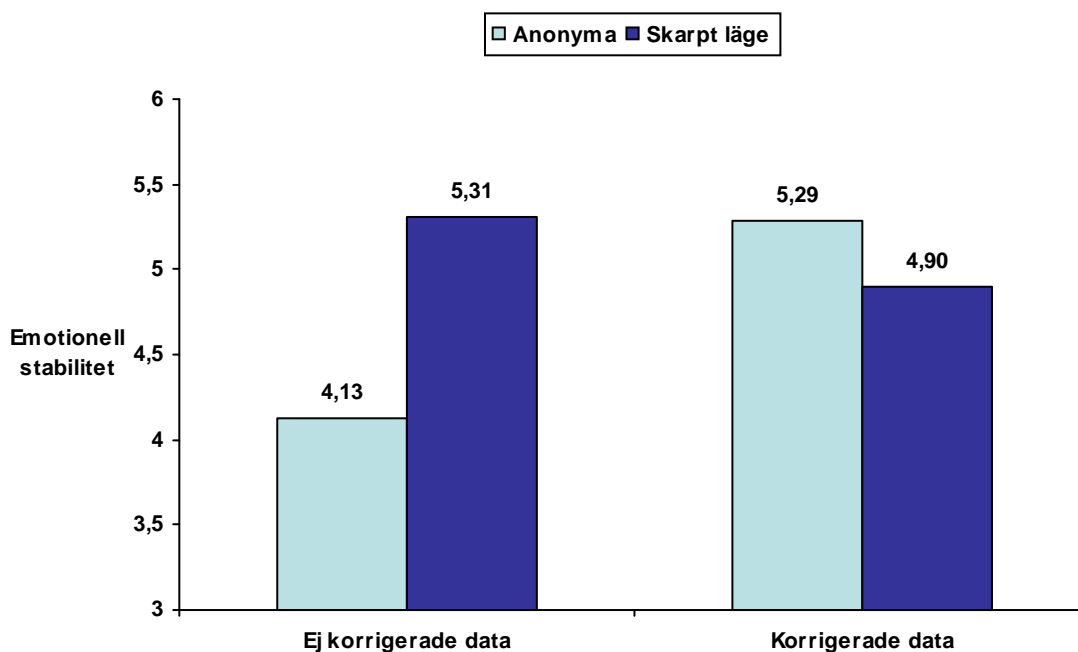
$r_{12}$  är korrelationen mellan overt och kovert skönmålning.

Proceduren kan givetvis utföras i statistikprogram som SPSS men den kan också programmeras explicit med hjälp av ovanstående formler. Denna metodik utnyttjar



skönmåningsskalan maximalt. Effekten av ett starkt incitament att svara taktiskt försvinner nästan helt, i gruppdatas genomsnitt, med denna typ av korrektion (Sjöberg, 2008; Sjöberg & Engelberg, 2005).

Den här tekniken har fungerat utmärkt och validerats i en studie av sökande till Handelshögskolan i Stockholm, vars värden jämfördes med studerande på skolan som tog samma test men i medvetande om att deras testvärden inte skulle ha någon för dem viktig betydelse (Engelberg & Sjöberg, 2005). Se Figur A2 som visar att skillnaden i okorrigerade värden starkt reduceras när vi genomfört korrektionen för skönmåning.

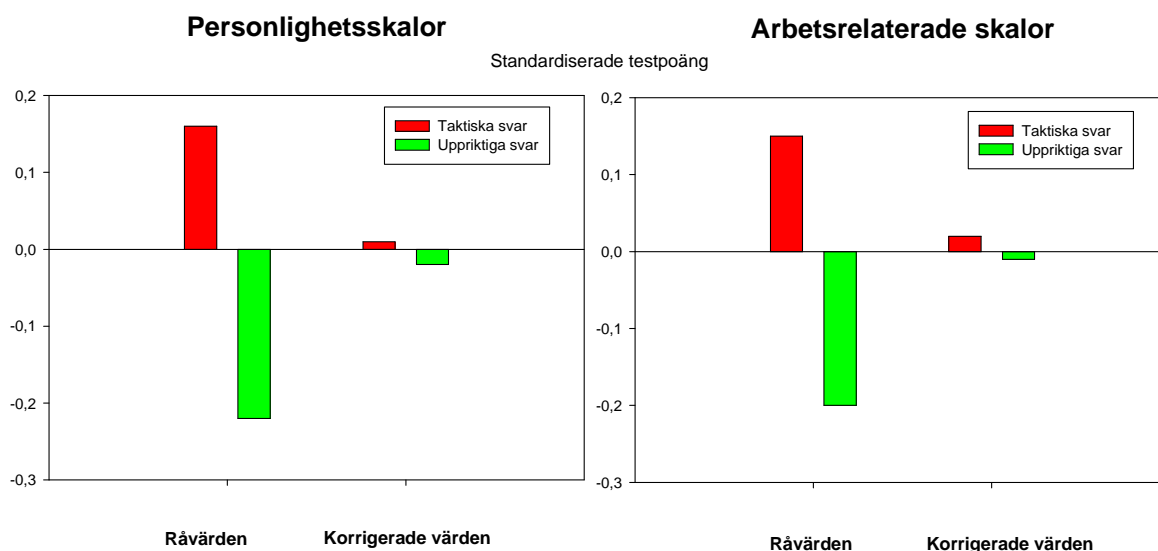


**Figur A2.** Effekten av korrektion för skönmåning vid jämförelse mellan anonym testning och skarpt läge, emotionell stabilitet. Medelvärde=5, standardavvikelsen=2.

Figur A3 visar resultatet av ett experiment som beskrivs närmare i den tekniska rapporten (Sjöberg, 2011 a). Somliga av de testade personerna uppmanades att skönmåla så mycket de kunde, andra uppmanades att ge ärliga svar. De som skulle bluffa uppmanades så här:

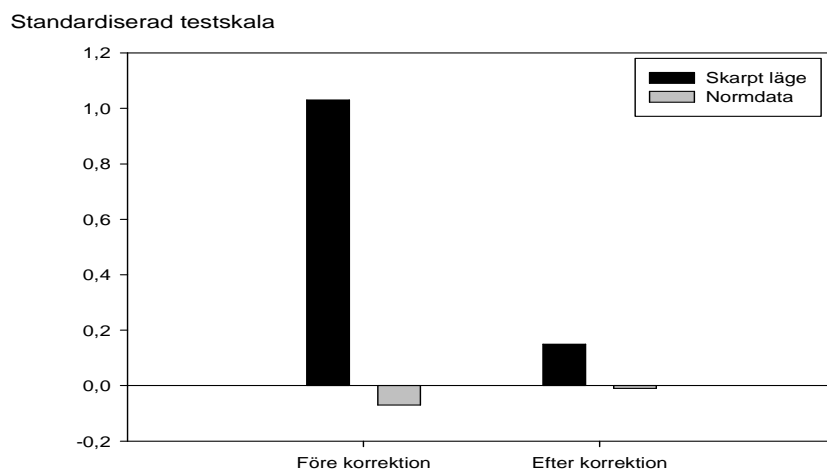
”Tänk på ett jobb som du mycket gärna skulle vilja ha. Tänk dig nu att du sökt just det jobbet och att den här testningen är ett mycket viktigt led i anställningsförfarandet. Svara på testuppgifterna så att du framstår som en person som är precis sådan som man söker till jobbet, men ändå trovärdig. Det kanske innebär att du bluffar en del, men det är just det som vi har för avsikt att studera i den här undersökningen. Känn dig därför fri att svara taktiskt!”

Skillnaden var stor före korrektion men försvann till 90-95 % efter korrektion, se Figur A3.



**Figur A3.** *Genomsnittliga effekter av korrektion för skönmålning, standardiserade skalor.*

Delar av *UPP*-testet (screening-modulen) används i rekryteringsplattformen *Katapult*<sup>15</sup>. De som där tar testet gör det i skarpt läge, dvs. i samband med att de söker ett jobb. Se Figur A4 för skillnader mellan testade i skarpt läge och normdatabasen (oskarpt läge), före och efter korrektion för skönmålning. Figuren gäller utåtvändhet men liknande resultat erhöles för andra testvariabler. Nästan hela (90 %) av effekten av skönmålning eliminerades. Detta resultat betyder också att normdatabasen är relevant även vid arbete i skarpt läge, om man nämligen korrigerar på det sätt som görs i *UPP*-testet.



**Figur A4.** *Utåtvändhet i skarpt läge, före och efter korrektion för skönmålning.*

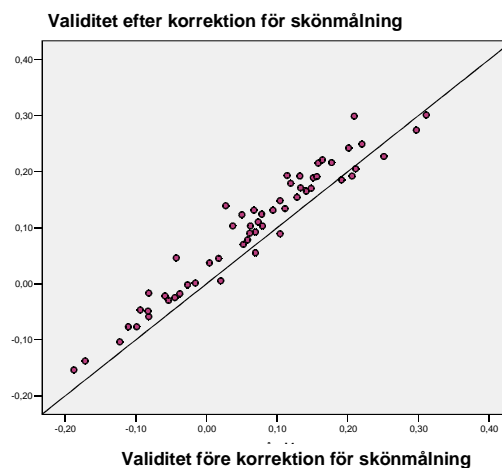
En korrektion av denna typ har stor praktisk betydelse. På individnivå kan effekterna bli mycket stora. Personer som hamnat i tätgruppen efter testning kan förpassas till botten av fördelningen, om de har skönmålat starkt. Kvinnor tenderar vid chefsrekry-

<sup>15</sup> [www.katapult.com](http://www.katapult.com)

tering att gynnas av att man genomför en korrektion (Sjöberg, 2005b), och att missgynnas om man inte gör det.

Enligt klassisk testteori bör de korrigerade testvärdena ha lägre reliabilitet än de ursprungliga. Detta beror på att de har formen av differensvärden. Analyser av samtliga indexvariabler i testet visade som väntat att det blev en viss sänkning av reliabiliteten efter korrektionen för skönmålning, från ett genomsnitt på 0,77 till 0,72. Å andra sidan minskade också felspridningen, dvs. precisionen i diagnostiska bedömningar av enskilda personer. Denna minskning uppskattades till, i genomsnitt, 14 %. För praktiska syften vid användningen av testet finns det alltså två skäl att korrigera för skönmålning: (1) Testvärdena ger efter korrektion en sannare bild av testpersonen eftersom eventuella försök att svara taktiskt i stort sett inte påverkar dem, och (2) Testvärdena är mindre påverkade av slumpfel i mätningarna.

Korrektionen för skönmålning ska enligt teorin bakom den leda till ökad validitet. Att en sådan effekt verkligen uppnås har nyligen påvisats i en studie av ledarskap (Sjöberg, et al., 2011). Fig. A5 illustrerar effekten; där har vi plottat resultaten från analys av medarbetares bedömningar av chefernas effektivitet i skilda avseenden. På x-axeln återges validiteten för varje skala och kriterium för okorrigerade data; på y-axeln motsvarande för korrigerade data.



**Figur A5.** *Validitet hos okorrigerade och korrigerade testskalor, medarbetares bedömningar som kriterium. Varje punkt motsvarar en testskala och en kriterieskala*

*UPP-testet* använder alltid korrektion för skönmålning enligt den metod som beskrivs i detta appendix.

## Appendix 2

Begåvningsstestet MHS-provet är en hierarkisk modell och har skapats för att differentiera i en högbegåvad population, det vill säga de individer som vid mönstring som hade bättre än medelbegåvning på Inskrivningsprovet (5 Stanine-poäng eller mer). MHS-provet mäter tre övergripande faktorer: Generell intelligens (G), som utgör toppen av hierarkin; samt Spatial förmåga (Gv) och Kristalliserad intelligens (Gc) som utgör den andra nivån uppifrån. G utgör således de testades sammantagna intelligens, medan Gv och Gc mäter de testades förmåga att mentalt bearbeta bilders information respektive förmågan att använda sig av tillägnad kunskap. Den sistnämnda är i mycket beroende av verbal förmåga. Ur ett militär perspektiv tror man att G och Gc är grundläggande för problemlösning, ordergivning och orderutförande samt beaktandet av olika tänkbara händelseutvecklingar. Gv tror man har betydelse för förmågan att förstå taktik, sätta samman enheter, samt förmågan att hålla koll på egna och fiendens trupper (Carlstedt & Widén, 1997).

De olika testdelarna redovisas i Tabell A2 och är hämtade från manualen för Militärhögskoleprovet av Carlstedt och Widén (1999).

**Tabell A2. Delskalorna i MHS-provet.**

Bongard	Denna del består av två grupper (Höger och Vänster) med vardera sex grafiska bilder. Varje grupp enas av ett mönster som samtidigt skiljer den från den andra gruppen. Bildgrupperna visas 20 sek. Därefter visas tre nya bilder i 10-30 sek beroende av svårighetsgrad. Uppgiften är att identifiera mönstret i vardera grupp genom att ange om de tre efterföljande bilderna tillhör "Vänster", "Höger", eller "Ingendera". Det är 24 uppgifter av bildmönster som skall identifieras. Alfa = 0,76.
Regler	Den här delen består av 22 uppgifter av ett svart och vitt rutmönster, där uppgiften är att upptäcka vilken princip som styr hur de vita rutorna placerar sig i de svarta rutorna. Alfa = 0,71.
Serier	Denna del består av 24 ofullständiga nummer serier, där uppgiften är att identifiera hur serien utvecklas och på så sätt kunna ange det nummer som logiskt sett skall följa den angivna serien. Alfa = 0,75.
Former	Detta test består av 2-4 svarta, tvådimensionella geometriska figurer med märkta sidor (bokstäver). Testtagaren skall under loppet av 15-30 sek mentalt sätta samman figurerna så att sidorna/bokstäverna från de olika figurerna korresponderar. Därefter skall denne ange vilken av fem figurer som motsvarar resultatet av roteringen. Totalt sett är det 22 figurer som skall sättas samman. Alfa = 0,70.
Prickar	Denna del består av 26 uppgifter och varje uppgift består av tre delar. Först presenteras ett enkelt mönster av 5x5 prickar där vissa prickar är svarta för att bilda ett mönster, därefter presenteras en ny kvadrat med 5x5 prickar, men nu är inga svartmarkerade och två pilar anger dels hur mycket det tidigare mönstret skall roteras, dels hur och vilka svarta prickar i det tidigare mönstret som skall flyttas. Testtagaren skall därefter identifiera det mönstret som den mentala roteringen resulterade i bland fem olika svarsalternativ. Alfa = 0,80.
Rotering	Den här delen består av 20 uppgifter och varje uppgift består av att identifiera om något av tre tredimensionella roterande objekt är lika eller olika en uppvisad förlaga. Alfa = 0,73.
Mönster	Testtagaren presenteras här med ett mönster i fem sekunder. Därefter presteras fem nya mönster och uppgiften består i att ange vilket av de fem svarsalternativ som motsvarade det första presenterade mönstret. Totalt är det 24 mönster som presenteras. Alfa = 0,59.
Ordkunskap	Ordkunskap testet är indelat i 80 uppgifter där testtagaren skall identifiera vilket av fem ord som har samma, eller nästan samma, betydelse som ett angivet ord. Orden är hämtade från fyra olika områden: Litteratur (Alfa = 0,67), Vetenskap (Alfa = 0,62), Teknik (Alfa = 0,66) och Aktuell debatt (Alfa = 0,75); med vardera 20 ord. Testet ger utöver måtten för de fyra områdena också ett övergripande mått (Alfa = 0,86) samt en indikation på intresseinriktning.
Associationer	I den här delen av ordkunskapstestet skall testtagaren hitta en förstavelse som passar ihop med de tre angivna ord som presenteras. Testet består av 26 uppgifter. Alfa = 0,64.