



PSYKOLOGISK METOD AB



***UPP*-testet**
Understanding Personal Potential
Teknisk manual

2010-06-20

Psykologisk Metod L Sjöberg AB arbetar med utveckling och användning av psykologiska test samt undersökningar av attityder och riskuppfattningar, andra psykologiska utredningar och tillämpad forskning.

Vår affärsidé är att bedriva arbetet i nära anslutning till den aktuella forskningen inom psykologin.

Skrifter utgivna av Psykologisk Metod AB

- Sjöberg, L. Bortom Big Five: Konstruktion och validering av ett personlighetstest. Rapport 2008:1.
- Sjöberg, L., & Möller, K. (2009). Sociala arbetsfunktioner och personlighet. Rapport 2009:1.
- Sjöberg, L. (2009). *UPP*-testet: Kriterierelaterad validitet. Rapport 2009:2.
- Sjöberg, L. (2009). *UPP*-testet: Korrektion för skönmålning. Rapport 2009:3.
- Sjöberg, L. (2010). *UPP*-testet: Tredje generationens personlighetstest. Rapport 2010: 1.
- Sjöberg, L.(2010). *UPP*-testet. Teknisk manual, Juni 2010 .
- Sjöberg, L (2010). *UPP*-testet: Mångfald gynnas av korrektion för skönmålning. Rapport 2010: 2.
- Sjöberg, L. (2010). A third generation personality test. Rapport 2010:3.
- Sjöberg, L. (2010). Emotionell intelligens och social förmåga hos ungdomar. Rapport 2010:4.
- Sjöberg, L. (2010). Faktorstrukturen hos *UPP*-testet. Rapport 2010:5.
- Sjöberg, L. (2010). *UPP*-testet och kundservice: Kriteriestudie. Rapport 2010:6.
- Sjöberg, L. (2010). *UPP/Screen*: Ett screeningtest av personlighet och begåvning. Rapport 2010:7.
- Sjöberg, L. (2010). *UPP*-testet: Användarhandbok. Juni 2010.

Innehåll

Innehåll.....	1
Förord.....	3
1. Inledning.....	4
Personlighetstestning.....	4
Testens syfte: förutsägelse och förståelse.....	5
2. Relation till Big Five-test (FFM-test).....	7
Traditionell FFM-mätning.....	7
Variabler i UPP-testet utöver FFM-dimensionerna.....	7
Övergripande personlighetsdimensioner.....	12
3. Räcker det med Big Five?.....	14
4. Utvecklingsarbetet.....	14
5. Uppgiftskonstruktion och innehållsvaliditet.....	16
6. Skalornas egenskaper: reliabilitet/homogenitet.....	16
7. Skalornas egenskaper: stabilitet.....	17
8. Skönmålning.....	19
9. Testdatas kvalitet.....	26
Kvalitetsmätning.....	26
Stämningsläge.....	29
Attityd till testet: ”Face validity”.....	32
10. Begreppsvalideringar.....	33
Personlighetsvariabler.....	34
Proxy-kriterier (arbetsmotivation, balans och andra arbetsrelaterade variabler).....	36
11. Validering mot externa kriterier och proxykriterier.....	37
Chefsansvar.....	37
Chefers arbetsresultat.....	39
Kundservice.....	40
Social anpassning i arbetsmiljön.....	51
Arbetsprestation (självbedömd).....	53
Arbetsmotivation och övriga arbetsrelaterade variabler (proxy-kriterier).....	55
Livsframgång.....	57
Sammanfattning och slutsatser om validering.....	58
12. Intellektuell förmåga.....	59
13. Narrativt utlåtande.....	61
14. Jämställdhet och mångfald: rättvisaspekter.....	62
Jämställdhet.....	63
Mångfald.....	65
15. Normativt eller ipsativt svarsformat?.....	71
16. Jasägartendensen i betydelsen globalt jasägande.....	71
17. Varför bättre än traditionella test?.....	72
18. Screeningmodulen.....	73
19. Källor till fördjupade kunskaper.....	80
20. Referenser.....	81

Bilaga 1. Några begrepp inom testteorin.....	88
Bilaga 2. Kompletterande information om begreppsvalideringar.....	90
Social förmåga och emotionell intelligens.....	90
Kreativitet.....	96
Stresskänslighet.....	97
Bilaga 3. Metodsynpunkter på validering.....	99
Stickprovets storlek.....	99
Externa kriterier.....	100
Korrelationer och andra sambandsmått.....	100
Mätfelens inverkan på korrelationer.....	101
Summering av ett tests validitet i fallet med flera underordnade testskalor eller flera kriterier.....	102
Slutsatser.....	102
Bilaga 4. Kompletterande information om reliabilitet.....	103

Förord

Denna manual ersätter tidigare upplagor av manualen för *UPP*-testet. Manualen bygger på den tekniska rapport som beskriver det forsknings- och utvecklingsarbete som ligger bakom konstruktionen och utvärderingen av testet [103]. Den kan beställas via hemsidan www.psykologisk-metod.se.

Manualen i föreliggande version innehåller nyheter som avspeglar det forsknings- och utvecklingsarbete som gjorts under 2009 och 2010: mätning av stämningsläge, mätning av nya övergripande dimensioner kallade stresskänslighet och chefspotential samt ingående analys av och rapportering om testdatas kvalitet för varje enskild testad person. Även testets ”face validity” mäts nu. Resultatprofilen har getts en grafisk utformning. Texterna i den narrativa rapporten har genomgått omfattande revision och finslipning.

Vi kommer fortlöpande att rapportera nya valideringsstudier och även annan forskning om *UPP*. Rapporterna kan beställas via hemsidan www.psykologisk-metod.se. Psykologisk Metod erbjuder även en utbildning för testanvändare som leder till certifiering.

Manualen är upplagd med tanke på att testet granskas av STP (Stiftelsen för tillämpad psykologi). STP tillämpat EFPA:s kriterier för testgranskning [59]. Dessa kriterier är svårtolkade. Vissa kriterier diskuteras i en särskild bilaga (Bilaga 3).

Frågor kan ställas till info@psykologisk-metod.se

1. Inledning

Syftet med denna manual är att ge en inledning till *UPP*-testet. Praktiska anvisningar för testets användning finns i en särskild användarhandbok. Denna manual innehåller introducerande text om personlighetstestning, en beskrivning av uppbyggnaden av testet och den forskning som utförts, samt principer för narrativa rapporter och övrig användning av testet i HR-arbete. Normdatamaterialet beskrivs. Det mesta av denna information, i större detalj, återfinns i den tekniska rapporten [103]¹, men viktig ny information finns i denna manual och i företagets rapportserie. I Bilaga 1 ges en kortfattad och elementär beskrivning av några grundläggande begrepp inom testteorin.

Personlighetstestning

Många är övertygade om att personligheten är en viktig faktor i arbetslivet. Den roll som psykologiska test spelar i sammanhanget diskuteras i [99]. En omfattande och aktuell diskussion av personlighetstestning i arbetslivet finns i [92].

En översikt av forskningen har visat att personligheten har stor betydelse för framgång i arbetslivet [87]. Det anförda arbetet är en s.k. meta-analys, en noggrant genomförd och kvantitativ sammanställning av ett stort antal undersökningar. Roberts et al. fann att personlighetens betydelse var jämförbar med betydelsen hos intellektuell förmåga, och klart viktigare än social bakgrund och kontaktnät. Exempel på yrkesroller där personligheten troligen spelar stor roll är chefer och säljare, men många andra kunde nämnas, kanske alla.

Hur kan man skaffa sig en uppfattning om någons personlighet? Och vad är egentligen personlighet? Den sista frågan brukar besvara med att personligheten kommer till uttryck i en persons *unika* beteende, tankar och känslor. Det är alltså variationen mellan människor som det handlar om. I psykologin talar man om differentiell psykologi, den del av ämnet som handlar om skillnader mellan människor. Ekmans grundläggande bok i differentiell psykologi [27] kan fortfarande rekommenderas. En modern framställning ges i [65].

Test är en av de viktigaste metoderna för att skaffa sig information om en person. De flesta personlighetstest som används inom arbetslivet är numera självrapporttest, och bygger på egenskapsteori. Man arbetar alltså med personlighetsegenskaper; exempel är utåtvändhet och emotionell stabilitet. Test av personlighetsegenskaper kan analyseras med en arsenal av väl genomarbetade psykometriska och statistiska metoder. Det finns emellertid också projektiva personlighetstest som ibland kommer till användning. De skall inte behandlas här, och åsikterna är mycket varierande om deras värde [40]. Självrapporttest har en etablerad validitet men hur stor den är påverkas av hur fokuserade testskalorna är.

Den egenskapsbaserade personlighetspsykologin är idag starkare än någonsin [65], och kunskaperna på området har ökat enormt under de senaste decennierna. Faktoranalytisk forskning ledde till en början till en allt vildare flora av personlighetsegenskaper som tycktes nästan omöjlig att överblicka, men från omkring 1990 har en stor förenkling åstadkommit genom femfaktormodellen (FFM). Den säger att det finns endast fem övergripande personlighetsfaktorer:

¹ Siffror inom hakparenteser hänvisar till referenslistan

- Utåtvändhet
- Emotionell stabilitet
- Noggrannhet
- Öppenhet
- Vänlighet²

Det finns en mycket omfattande forskning om FFM och många studier har visat att denna övergripande struktur är robust och kan replikeras på många språk och i skilda kulturer. Test som NEO-PI-R och HPI bygger på denna begreppsram (även om HPI testar 7 och inte 5 övergripande dimensioner; Öppenhet och Utåtvändhet har i HPI splittrats i vardera två dimensioner).

FFM-modellen har inte gett uppmuntrande resultat när det gäller validitet i arbetslivet. För ett exempel, se aktuella resultat med OPQ-testet som snarast ligger under den traditionella testvaliditeten från 1960-talet [11]. Meta-analyser av FFM-test har gett mycket svaga resultat [72-73].

Behövs det då ett nytt test? Ja, av tre skäl:

1. Befintliga test har fortfarande inte bättre prognoskraft till viktiga dimensioner i arbetslivet (kriterier) än vad personlighetsvariabler hade när Mischel skrev sin berömda uppgörelse med personlighetspsykologin år 1968 [71]. Detta gäller även FFM-dimensioner. Som visas i den tekniska rapporten finns två troliga skäl till dessa nackdelar. Dels är de övergripande FFM-dimensionerna alltför generella i förhållande till arbetslivsförhållanden. De är abstrakta och därför mindre relevanta [94; 98]. Dels är de alltför kognitivt inriktade och arbetar med ett utifrånsperspektiv; personen förväntas bedöma sig själv ”utifrån” och det frågas inte efter känslor. Men just känslor är mycket viktiga i många sammanhang, inte minst inom arbetslivet.
2. Befintliga test har inga effektiva metoder för att korrigera för skönmålning. Många testade personer svarar taktiskt och försöker framställa sig själva i en alltför positiv dager. Det påstås ibland att detta inte är något stort problem vid testning med självrapporttest. Detta är ett felaktigt påstående.
3. Nästan alla test på den svenska marknaden är översättningar av amerikanska, brittiska eller tyska test. Det är inte troligt att test är oberoende av sitt kulturella sammanhang. Det är angeläget att svensk, oberoende forskning kommer till stånd som kan lägga grunden för test som passar vår kultur. Valideringar av amerikanska test i USA kan inte utan vidare antas gälla också för svenska förhållanden. Översättningar till svenska kan vara mycket svåra att göra på ett helt tillfredsställande sätt, och varje ord måste vägas på guldväg när det gäller testuppgifter. Se t ex not 2.

Testens syfte: förutsägelse och förståelse

Den som studerar teorin för psykologiska test märker snart att det handlar om att göra prognoser. Test utvärderas efter hur bra de är i den funktionen, oftast genom att man skaffar in kriteriedata, t ex data på arbetsprestation, och sedan beräknar sambandet mellan test- och kriteriedata. Detta brukar kallas prognostisk eller samtidig validering av testet. Det diskuteras

² Den engelska termen är ”agreeableness”. ”Agreeable” betyder trevlig eller angenäm och det är ju inte riktigt detsamma som vänlig.

ganska sällan om detta verkligen är lämpligt som en yttersta mätare av testens värde. Ändå finns det ett flertal goda skäl att reflektera över den saken. Det ska göras här med utgångspunkt i förhållandena för personlighetstest men liknande resonemang kan föras även för färdighetstest.

För det första: framtida resultat, t ex i form av arbetsprestation, beror alltid på många faktorer utöver personligheten hos den som testats. Det kan gälla individanknutna aspekter som sjukdomar eller sådant som handlar om individen i förhållande till sin omgivning. Det handlar alltså om externa faktorer som ligger utanför personligheten och som kan vara av mycket stor betydelse. Det är i princip omöjligt att förutsäga alla dessa faktorer utifrån enbart data på personlighet, och de sätter ett tak för hur mycket som är möjligt att förutsäga från personlighetstest. Vi vet sällan eller aldrig var detta tak är beläget. Om ett test kan sägas vara bra eller dåligt utifrån resultatet av en validering beror på i vilken grad det alls är möjligt att göra prognoser. Det finns ingen forskning, såvitt bekant, där man försökt uppskatta hur stor betydelse faktorer utanför personligheten har (med undantag för intellektuell förmåga).

För det andra: det finns oerhört många tänkbara ”kriterier”, och alla har de olika samband med testen. Valet av ett visst eller några få kriterier i syfte att utvärdera värdet av ett test är tämligen godtyckligt.

För det tredje: användningen av personlighetstest är sällan *i praktiken* inriktad på att göra specifika prognoser. Vad som eftersträvas är i stället *förståelse* av personligheten. *UPP*-testet arbetar med den filosofin och heter därför ”Understanding Personal Potential”. Hur bra ett test är för att skapa förståelse kan inte utvärderas med enkla psykometriska och statistiska beräkningar. Det beror ju bl. a på *vems* förståelse det handlar om: den testade personens egen förståelse (självkänedom), eller rådgivares, konsulter och beslutsfattares förståelse. Det kan tyckas att vi kommer ut på ett gungfly av subjektiva omdömen här. Det verkar onekligen tryggare att begränsa sig till prognostisk eller samtidig validitet. Den gamla positivistiska psykologin såg det som en självklarhet att arbeta på det sättet. *UPP*-testets namn är ”Understanding Personal Potential”, vilket vi tror är just vad som efterfrågas av användare. Prediktion är också viktigt, och ännu ett av *UPP*-testets syften, men inte det mest centrala i testanvändning.

För att ge en begriplig rapport till testanvändaren och den som tagit testet är det lämpligt att summera testskalorna med staninepoäng, alltså i nio steg och efter transformering till normalfördelning. Flera steg än nio är olämpligt eftersom det ställer för stora krav på mänsklig informationsbehandling. Särskilt olämpligt är det troligen att använda percentiler (100 steg) som dels har för många steg, dels är ett begrepp som är svårt att förstå för dem som inte är särskilt utbildade i statistik [10].

2. Relation till Big Five-test (FFM-test)

Traditionell FFM-mätning

Det finns två filosofier bakom mätning av FFM-dimensionerna.

1. Test som HPI och NEO-PI-R arbetar med ett stort antal delskalor [kallade ”facets” i NEO-PI-R och ”homogeneous item composites” (HCI) i HPI]. Antalet delskalor är 30 i NEO-PI-R (totalt 240 testuppgifter) och 41³ i HPI (totalt 206 uppgifter). Problemet är emellertid att dessa skalor i åtskilliga fall bara har en vag relation till det övergripande begrepp som de avser att mäta. De fem övergripande begreppen blir samlingsnamn för mycket heterogena innehåll. Faktoranalyser har visat att det *finns* gemensamma komponenter hos delskalorna som kan, med lite fantasi, tolkas som begreppen i FFM, men mycket av den intressanta psykologin och viktiga informationen bakom delskalorna går förlorad när de används enbart som underlag för FFM-dimensionerna.
2. Ett annat sätt att arbeta är att ta fram fokuserade testuppgifter som innehållsmässigt nära ansluter till FFM-dimensionerna. Så arbetar man inom den internationella databasen IPIP, se länken <http://ipip.ori.org/ipip/>. Här finns utmärkta möjligheter att efter översättning skapa korta och effektiva test för mätning av dimensionerna i FFM. Femtio uppgifter räcker väl, 10 för var och en av dimensionerna. Omfattande test av typ NEO-PI-R och HPI ger inte mätningar av FFM-dimensioner som ger mera information än ett kortfattat test av IPIP-typ (även om delskalor kan göra det).

UPP-testet mäter FFM-dimensionerna enligt samma ansats som IPIP-metodiken, men med nykonstruerade uppgifter, inte översättningar. De skalor som konstruerats för *UPP*-testet har relaterats till IPIP-skalor för begreppsvalidering. Dessutom tillkommer ett antal övriga dimensioner, varav somliga, men inte alla, kan återfinnas bland delaspekterna i NEO-PI-R och HPI. Dessa har validerats på skilda sätt, med resultat som beskrivs i denna manual, i den tekniska rapporten och i Psykologisk Metods rapportserie.

Variabler i UPP-testet utöver FFM-dimensionerna

Kreativitet

Kreativitet är en sedan länge ganska försummad dimension, men det finns goda skäl att uppmärksamma den mera i arbetslivet [93; 102]. Kreativitet hos chefer är en viktig aspekt på ledarskap och kan mätas med psykologiska test [91]. Allmänt sett har forskningen om individuella skillnader i kreativitet lett till positiva slutsatser om Guilfords ursprungliga begrepp divergent tänkande [36] och om öppenhet och liknande personlighetsvariabler [123].

³ Med så många delskalor och relativt få testuppgifter blir med nödvändighet många av delskalorna otillräckligt reliabla.

Social förmåga

En intressant dimension är *social förmåga* som knyter an till begreppet social intelligens [56; 122]. De tidiga försöken att mäta social intelligens avbröts därför att sambandet med verbal intelligens var högt – detta framstår numera inte som välbetänkt. Dimensionen är viktig, och kan mätas, oberoende av verbal intelligens. Se avsnitt 12 där data rapporteras som visar på nollsamband mellan verbal och annan intellektuell förmåga å ena sidan, social förmåga å den andra. I avsnitt 19 rapporteras data från användning av UPP-testets screeningmodul som visar på ett lågt samband ($r = 0,20$) mellan verbal intelligens och samarbetsvilja, ett begrepp som är besläktat med social förmåga.

Positiv grundattityd

Ytterligare en dimension har kallats *positiv grundattityd* [114]. Den dimensionen mäter grundläggande positiv inställning till världen, människorna och den egna personen. En person med högt värde i den dimensionen letar inte efter fel, ser ljusa sidor hos tillvaron och glömmer och förlåter kränkningar och misstag. I en översiktsartikel skriver Lyubomirsky el al. [60]:

“Numerous studies show that happy individuals are successful across multiple life domains, including marriage, friendship, income, work performance, and health”. (Sid. 803)

Det fanns alltså goda skäl att inkludera denna dimension. I vår egen forskning hade vi tidigare fått stöd för den [114].

Självförtroende

Självförtroende är en viktig aspekt, som varit föremål för mycket omfattande forskning [39] och har nära anknytning till Banduras [3-4] teori om den egna upplevda förmågan (self-efficacy). I arbetslivet är det en viktig variabel [52]. I förhållande till arbetsprestation har självförtroende ett samband omkring 0,40 enligt meta-analyser [15; 118]. Det finns en skillnad mellan självuppskattning (self esteem) och självförtroende, men det två är nära relaterade [119]. I denna manual kommer inte den distinktionen att göras.

Uthållighet

Uthållighet är ytterligare en viktig dimension som är relaterad till självförtroende [12] och till emotionell intelligens. Dimensionen har stor betydelse i arbetslivet inte minst för säljare [13], som ofta misslyckas, det ligger ju i försäljningsarbetets natur. Misslyckande är oftast en destruktiv upplevelse som leder till fortsatta misslyckanden [117] eller att man inte försöker på nytt. Ytterst få personer har försökt sig på att göra *mycket* svåra saker⁴. Man ger upp efter att ha misslyckats ett par gånger. ”Inlärd hjälplöshet” kan bli resultatet av en serie misslyckanden. De som lyckas i arbetslivet gör det delvis för att de inte ger upp och resignerar inför svåra uppgifter. Det är oftast framgång med något riktigt svårt som leder till stora belöningar. Uthållighet är en mycket viktig egenskap i arbetslivet.

⁴ Ännu opublicerad forskning.

Perfektionism

Noggrannhet är en av grunddimensionerna i FFM, men den kan gå till överdrift och bli till en tvångsmässig upptagenhet med detaljer. Dimensionen *perfektionism* mäts därför i testet. Perfektionism är en variabel som ofta nämns som en negativ prognosfaktor, och det finns ganska omfattande psykometriska resultat [86] om den saken. Noggrannhet är normalt en positiv egenskap. I *UPP*-testet mäts perfektionism efter psykometrisk korrektion för noggrannhet. Sambandet mellan noggrannhet och perfektionism är långt ifrån perfekt.

Emotionell intelligens

Emotionell intelligens (EI) är en ofta diskuterad dimension, se [51; 97]. Med EI menas förmågan att identifiera och hantera emotioner hos sig själv och andra. EI mäts i *UPP*-testet dels med en självrapportskala, dels med prestationsuppgifter som ställer krav på den testade personen att identifiera vilka emotioner som uttrycks i fotografier av ansiktsuttryck.

En mycket omfattande och intressant genomgång av forskningen om emotionell intelligens (EI) har publicerats av Joseph och Newman [51]. De har gjort en sammanställning av resultat med prestationsmätning av EI, och självrapportformulär. De anser som så många andra forskare att prestationsmätning är den teoretiskt mest tillfredsställande metodiken, eftersom det rör sig om en sorts ”intelligens”. Självrapportmetoder är de mycket kritiska till på samma teoretiska grund. Sambandet mellan prestationsmått och självrapportmått är nästan noll, i metaanalysen uppskattat till en korrelation på 0.12. Om man går vidare till validering mot arbetsprestation visar det sig emellertid att prestationsmått har ett mycket svagt samband med sådana kriterier, och inte ger något utöver kognitiva test och det traditionella Big Five, medan självrapportskalor ger ett visst tillskott till prognosförmågan. Aktuell forskning visar samma sak [18].

Joseph och Newman skiljer mellan två typer av självrapportformulär: sådana som har ett mycket heterogent innehåll av typ ”lite av varje” (mixed models – typexempel finns i många allmänt använda EI-test som Bar-Ons EQi-test) och sådana som mäter den självbedömda förmågan att klara EI-uppgifter. Bara den senare typen av självrapporter har ett tydligt samband med kriterier och något att ge utöver annan information. Sambanden är särskilt starka för arbetsuppgifter som är emotionellt påfrestande, t ex i serviceyrken där det gäller att alltid vara ”glad och trevlig”.

Denna artikel är en milstolpe i litteraturen om EI. Slutsatserna bygger på hundratals empiriska studier som författarna noga analyserat. Den är nedslående för dem som tror på prestationsmått av typ MSCEIT [67], och den pekar framåt mot konstruktionen av nya och bättre typer av självrapportskalor. Prestationsmåttens dåliga egenskaper har ju varit kända ganska länge men många forskare har ändå hållit fast vid dem på grund av deras teoretiska fördel. Det är dags att tänka om.

Det kan noteras att Goleman, som med sin bok från 1995 [34] skapade en marknad för konsultverksamhet i EI-branschen, av Joseph och Newman återigen kritiseras för sina påståenden om att EI är av överlägset stor betydelse. Data motsäger honom på ett övertygande sätt, vilket inte betyder att EI helt saknar betydelse. Joseph och Newman visar ju att EI-skolor

kan ge ett värdefullt tillskott av information, om mätningen görs på rätt sätt och speciellt i emotionellt relevanta arbetssituationer.

Samarbetsvilja – passiv aggression

Passiv aggression är ett uttryck för social dysfunktion som i extrem form medför starka förseningar av arbetsuppgifter och ovilja att samarbeta konstruktivt med andra. I testet utvärderas denna skala omvänt och kallas då för *samarbetsvilja*.

Passiv aggression handlar kort uttryckt om personer som saboterar en verksamhet genom bristande lojalitet, dold fiendlighet och ovilja att samarbeta. Detta sker ofta helt eller delvis omedvetet. Det är en klinisk diagnos men dimensionen varierar troligen steglöst i populationen, och subkliniska varianter kan mycket väl förekomma utan att vara extremt utpräglade och uppenbara för alla. Diagnosen krävde enligt DSM-III-R att minst 5 av följande 9 beteenden förekom hos en person: skjuter upp arbetsuppgifter, motsätter sig uppgifter som han eller hon inte vill utföra, arbetar ineffektivt med sådana uppgifter, klagar oresonligt på krav, ”glömmer” åtaganden och skyldigheter, tror sig göra ett mycket bättre jobb än vad andra anser, tar inte emot förslag från andra på ett konstruktivt sätt, gör inte sin del av jobbet och kritiserar chefer på ett oskäligt och omotiverat sätt. I DSM-IV har denna personlighetsstörning satts på en ”observationslista” i avvaktan på fortsatt forskning.

Passiv aggression är troligen vanlig och kan ha stora negativa konsekvenser. Många har träffat på den i arbetslivet. Murphy och Oberlin [75] skriver:

”In the workplace, hidden anger is not merely annoying; it can have long-lasting, even deadly consequences. In can cost millions – sometimes billions – in dollars of lost contracts, reduced efficiency, and even worker safety in some settings”. (Sid. 36).

I litteraturen har beskrivits ett syndrom som består av undvikande, försenade eller inte alls utförda arbetsuppgifter, bristande lojalitet, hyperkritisk men dold inställning till chefer, irriterbarhet, missnöje med den egna personen och sociala anpassningsproblem. Millon [70] inkluderade negativism, och formulerade dimensionen på följande sätt:

8A - Negativistic (Passive-Aggressive)

8A.1 - Temperamentally Irritable

8A.2 - Expressively Resentful

8A.3 - Discontented Self-Image

Syndromet mäts även i Hogans “Development survey” [44]. Men varken Millon eller Hogan täcker in alla aspekter. Enligt en utmärkt sammanställning i Wikipedia ingår följande komponenter i passiv aggression:

- Ambiguity
- Avoiding responsibility by claiming forgetfulness
- Blaming others
- Chronic lateness and forgetfulness
- Complaining
- Does not express hostility or anger openly
- Fear of competition
- Fear of dependency

- Fear of intimacy
- Fears authority
- Fosters chaos
- Intentional inefficiency
- Making excuses and lying
- Obstructionism
- Procrastination
- Resentment
- Resists suggestions from others
- Sarcasm
- Sullenness

Detta är mycket allvarliga problembeteenden som torde vara ganska vanliga i arbetslivet [42; 68; 76; 124]. En populärt hållen svensk diskussion finns i [77].

Personer som ligger högt i passiv aggressivitet gör inga bra arbetsinsatser men det kan vara svårt att upptäcka dem i tid eftersom de kan vara charmerande och lova guld och gröna skogar. Det handlar om allvarlig dysfunktion, som inte förväntas hos en person som i övrigt har goda egenskaper och det tar sin tid innan omgivningen inser att det som tycktes vara tillfälliga misstag ingår i ett bestående mönster. Terapi kan hjälpa, men för det finns naturligtvis ingen garanti [55] för det. En del bra idéer om hur man kan handskas med passiv aggression finns på <http://www.coping.org/anger/passive.htm>.

Trettio testuppgifter för att mäta passiv aggression. Faktoranalys gav 4 korrelerade faktorer:

- Allmän social dysfunktion
- Avvisande av auktoriteter
- Döljande av emotioner
- Avvisande av socialt stöd

Kärnan i passiv aggression är alltså social dysfunktion och bristande samarbetsvilja. Denna dimension kallas därför samarbetsvilja, efter att den utvärderats så att höga värden på passiv aggression ges låga värden på samarbetsvilja, och tvärtom. Den är inkluderad i *UPP*-testet med en kortfattad, reliabel och validerad skala [108]. I ett material om 3586 testade (jobbsökande) personer [111] verkar problem med passiv aggression finnas i ca 10 % av fallen. En klusteranalys visade att 328 av de 3486 testade personerna tillhörde grupper som hade särskilt låg samarbetsvilja (drygt 9 %). Data var korrigerade för skönsmålning.

Varför finns det så lite psykometriskt arbete om passiv aggression? Ett skäl kan vara att syndromets status inom psykiatrin länge varit osäker [75]. Flera olika definitioner har förekommit i olika upplagor av DSM, och det är möjligt att forskarna har avskräckts av begreppets osäkra status. Detta hindrar inte att det finns en viktig realitet bakom denna långvariga och snåriga diskussion av definitionsfrågor, och att kärnan i begreppet väl låter sig fångas som ”bristande samarbetsvilja”. I ett aktuellt arbete finner man starkt stöd för begreppets giltighet [46]. Det är definitivt en aspekt som bör uppmärksammas mera i praktiskt arbete med urval och utveckling av personal.

Övergripande personlighetsdimensioner

UPP-testet mäter tre övergripande dimensioner, nämligen jagstyrka, stresskänslighet och ledarskapsförmåga eller chefpotential. Dessa dimensioner är användbara om man vill ha en övergripande bedömning på grundval som sammanfattar en stor del av de enskilda personlighetsdimensionerna i testet.

Jagstyrka, stresskänslighet och chefpotential korrelerar högt sinsemellan, men inte perfekt. De höga korrelationerna beror givetvis på att delvis överlappande variabler ingår i indexen. Alla tre variablerna används i testet eftersom de kan vara av särskilt intresse i speciella situationer. De aggregerade dimensionerna beskrivs i det följande.

Jagstyrka

Elva av delskalorna i testet avspeglar begreppet jagstyrka. Dessa var:

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

En jagstark person är realistisk, förstår väl både sig själv och andra, och är även långsiktig. Jagstyrka innebär en förmåga att behärska sina impulser att eftersträva kortsiktiga vinster och njutningar. En jagstark person är positiv till människorna utan att vara orealistisk, och sällan eller aldrig manipulativ. Som chef kan han eller hon delegera och ge ansvar till medarbetare, är inte defensiv utan kan ta till sig kritik. Den jagstarke ser med distans och humor på sig själv och är medveten om sina brister. Se [100].

De 11 delskalorna som avspeglar jagstyrka ingick i ett särskilt index för att mäta detta begrepp, reliabilitet = 0,72. Personlighetsskalorna uppvisade en endimensionell struktur, med undantag för noggrannhet och perfektionism som föll ut i en särskild faktor. Jagstyrkan uppskattades som medelvärdet av dessa personlighetsskalor, efter korrektion för skönmålning (se avsnitt 8).

Stresskänslighet

Högt värde i dimensionen stresskänslighet kan innebära problem för personen att fungera bra under stress: negativa emotioner som ilska, rädsla, eller nedstämdhet sänker livskvaliteten och stör arbetet. Under stress kan personen bli låst vid detaljer (perfektionism) och ”tunnelseende”, dvs. ha svårt att ta med alla viktiga aspekter i problemlösning och beslutsfattande. Uthålligheten kan bli nedsatt, och relationen till andra människor negativ.

Stresskänslighet 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

Reliabiliteten för risk under stress var 0,61, ett något lågt värde men dimensionen kan ändå vara till nytta.

Chefspotential

Följande variabler användes för att skapa den aggregerade variabeln chefpotential:

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

Reliabiliteten för denna variabel var 0.80.

Specialanpassade aggregerade variabler

Vid behov kan nya aggregerade variabler skapas t ex för att mäta lämplighet för en viss befattning. På så sätt får man en övergripande bedömning av testresultatet, vilket kan i hög

grad underlätta användningen av testet för praktiska syften. De nya aggregerade variablerna bygger på användarens egen bedömning av hur viktigt det är med goda testresultat i de olika dimensionerna (lämplighet) och/eller hur viktigt det är att undvika låga värden (olämplighet). Båda dessa ansatser kan vara av värde. Det är inte givet att de är varandras spegelbilder.

Eftersom testet utvärderas i förhållande till en normgrupp är det enkelt att ge normerade värden på de aggregerade skalorna för ”lämplighet” och olämplighet”. Se närmare användarhandboken.

3. Räcker det med Big Five?

De tillkommande 8 testvariablerna (uthållighet, samarbetsvilja etc.) användes som beroende variabler i multipla regressionsanalyser där de 5 skalorna som mäter femfaktormodellen användes som förklarande variabler, se Tabell 3.1 för resultaten. Samtliga data har först korrigerats för skönmålning.

Tabell 3.1. Andel av variansen som förklaras av FFM (”Big Five”) i de tillkommande skalorna.	
Skala	Förklarad varians
Uthållighet	0,259
Samarbetsvilja	0,202
Positiv grundhållning	0,054
Självförtroende	0,206
Social förmåga	0,333
Emotionell intelligens, självrapport	0,130
Emotionell intelligens, emotionsbedömningar	0,011
Kreativitet	0,264
Perfektionism	0,320
Medelvärde	0,198

Tabellen visar att större delen av variansen i de tillkommande variablerna inte kunde återföras på den traditionella FFM – modellen. Ca 80 % av variationen i de tillkommande variablerna tillförde något verkligt nytt utöver FFM. Att det inte var slumpfel som svarade för de låga nivåerna på förklarad varians visar Tabell 6.1 – reliabiliteterna hos samtliga variabler var på god nivå. Det räcker alltså inte med Big Five.

4. Utvecklingsarbetet

Arbetet med att utveckla *UPP*-testet har pågått under fem år. Under denna tid har frågeställningar och målsättningar utvecklats och fördjupats. I sin nuvarande form innehåller *UPP*-testet 5 FFM-skolor och därutöver 8 övriga personlighetsskolor, 2 skolor för att mäta skönmålning (en overt och en kovert) samt 6 skolor för mätning av arbetsrelaterade attityder och arbetsmotivation. Dessutom mäts 5 aspekter på datakvalitet⁵, som inte är till för att

⁵ Intra-individuell variation, jäsägandetendens, strukturlikhet, antal ej besvarade frågor samt ett övergripande mått på stämningsläge. De tre förstnämnda sammanvägs till ett aggregerat mått på datakvalitet.

beskriva den testade utan för att bilda underlag för beslut om testning. Testets utveckling och validering har beskrivits i 17 studier. Tabell 4.1 ger en överblick av dem.

Tabell 4.1. Översikt av studierna.

Studie nr. Tidpunkt för studien	Syfte(n)	Antal deltagare
1 Nov -05	Grundkonstruktion av testet, validering mot självbedömd arbetsprestation, begreppsvalidering av skalor för att mäta skönmålning, faktoranalytisk grund för jagstyrka, relationen till intellektuell kapacitet, relation till arbetsrelaterade attityder	159
2 Feb -06	Kontroll av reliabiliteter. Begreppsvalideringar av Extraversion, Social förmåga, Emotionell stabilitet, Noggrannhet, Uthållighet, Positiv grundhållning, Emotionell intelligens, Självförtroende och Passiv Aggression. Dessa skalors relation till arbetsrelaterade attityder, samt undersökning av tillskottet i validitet utöver FFM-dimensionernas förklaringskraft.	328
3 Okt -06	Begreppsvalideringar av Extraversion, Social förmåga, Emotionell stabilitet, Noggrannhet, Självförtroende, Passiv Aggression och Emotionell Intelligens. Stabilitet hos samtliga skalor. Korsvaliderade reliabiliteter. Dessa skalors relation till arbetsrelaterade attityder	56
4 Okt - 07	Begreppsvalideringar av Vänlighet, ny skala, Öppenhet, ny skala, Kreativ attityd, Perfektionism och jagstyrka. Stabilitet hos samtliga skalor. Korsvaliderade reliabiliteter. Dessa skalors relation till arbetsrelaterade attityder.	51
5+6 Maj - 06	Faktoranalys av Passiv Aggression. Demografiska samband med skalorna. Begreppsvalidering av Passiv aggression, Social förmåga, Emotionell intelligens, Positiv grundattityd, Arbetsvilja, Förändringsvilja och Resultatorientering. Analys av jagstyrka. Validering mot arbetsrelaterade attityder.	201+202
7 Maj - 07	Begreppsvalidering av Emotionell Intelligens och Social Förmåga	149
8 Tidigare insamlade data	Begreppsvalidering av Kreativ Attityd	45
9 Aug -08	Begreppsvalidering av Perfektionism, Arbetsintresse, Förändringsvilja, Uthållighet, och Balans.	57
10 Aug -08	Begreppsvalidering av skalorna som avser att mäta skönmålning	133
11 Maj -08	Normering, analyser av jasägartendens, skalornas relation till arbetsrelaterade attityder, samt undersökning av tillskottet i validitet utöver FFM-dimensionernas förklaringskraft, prövning av den normativa mätningens validitet, analys av jagstyrka. Slutlig bestämning av reliabiliteter.	823 (612 kompletta)
12 april 2009	Samband mellan personlighet och verbal/analytisk förmåga	120
13 maj 2009	Emotionell intelligens, social förmåga och social anpassning i arbetet, validitet mot externt kriterium [106; 115]	107
14 oktober 2009	Screeningmodulen använd i skarpt läge vid rekrytering [105]	950 ⁶
15 mars 2010	Begreppsvalidering av EI och social förmåga [107]	128
16 maj 2010	Kriterievalidering i kundservicearbete	53
17 förstudie	Chefers arbetsresultat och jagstyrka	66
Sammanlagt antal deltagare i samtliga studier		3628

I det följande redogörs för ett urval av resultaten. Den fullständiga redovisningen finns i den tekniska rapporten för studierna 1-11, för studie 12 i denna manual samt i särskilda rapporter för studierna 13 -16. Studie 17 beskrivs i [100].

⁶ Senare utökat till ca 5000 testade.

5. Uppgiftskonstruktion och innehållsvaliditet

Testkonstruktionen baserades delvis på ett tidigare test som använts med framgång [100], samt omfattande forskning om emotionell intelligens som började med antagningsprövningar vid Handelshögskolan i Stockholm [28; 97; 104]. För varje skala i testet skrevs betydligt flera uppgifter än vad som avsågs ingå i slutversionen. Urvalet baserades sedan på uppgifternas empiriska egenskaper, dvs. medelvärden och korrelationer med övriga uppgifter som skulle ingå in i skalan.

Vid konstruktionen av testuppgifterna eftersträvades en avtäckning av skilda innehållsaspekter som alla ingick i begreppets domän. Det var viktigt att en skala inte fick bli alltför snäv utan att den verkligen omfattade en stor del av domänen. Det var också angeläget att uppgifterna hade både positiv och negativ riktning, ungefär lika många. I enstaka fall var den målsättningen svår att uppnå.

Vad domänen för en testvariabel är och vad den innehåller kan inte *a priori* ges en exakt definition. Testuppgifter vars empiriska egenskaper tyder på att de inte passar in har uteslutits ur de slutliga skalorna och acceptabla reliabiliteter har uppnåtts. Värdena på reliabilitet har sedan kontrollerats i nya stickprov. Innehållsvaliditeten i skalorna får sin belysning genom det program för begreppsvalidering som redovisats, och även i validering mot externa kriterier. Vi har valt denna ansats i stället för att försöka kvantifiera innehållsvaliditeten på grundval av ett antal bedömares bedömning av enskilda testuppgifters relevans, en metod som ibland används [58]. Se senare avsnitt i denna manual och den tekniska rapporten.

6. Skalornas egenskaper: reliabilitet/homogenitet

Reliabiliteten har undersökts i samtliga studier men här ges endast data från Studie 11 som gjordes med det slutliga testet och med en stor grupp (823 deltagare). Tabell 6.1 visar homogenitetsvärdena och en del övrig information om delskalorna. Analyser av andra datamängder och har gett mycket likartade resultat, se Bilaga 4.

Den metodik som använts här är klassisk testteori som finns beskriven i många läroböcker, se t ex Gulliksen [37] eller Magnusson [63]. Den ursprungliga analysen av testuppgifterna finns beskriven i den tekniska rapporten.

Innehållsvaliditet (content validity) är en aspekt som ofta är relevant vid konstruktion av kunskapsprov. Man önskar att alla viktiga aspekter i ett visst kursinnehåll ska täckas av uppgifterna i kunskapsprovet. Det är mindre uppenbart hur innehållsvaliditet ska kunna uppskattas för uppgifter i ett personlighetstest. Skalorna i *UPP*-testet konstruerades genom att i en första fas skriva uppgifter som såvitt kunde bedömas täckte av alla viktiga aspekter på personlighetsegenskapen i fråga. Det är väsentligen en subjektiv bedömning att avgöra om så är fallet. Denna bedömning får sedan sitt stöd genom den psykometriska uppgiftsanalysen, varvid både alfa-värdet och den genomsnittliga interkorrelationen mellan uppgifterna beaktas. Vissa egenskaper, som emotionell intelligens och samarbetsvilja, tycktes definiera domäner som var mera löst sammanhållna än andra, som extraversion och positiv grundhållning. I de förstnämnda fallen krävdes därför längre skalor. Det slutliga ”beviset” för innehållsvaliditet är

att skalorna har förväntade relationer till andra variabler, och detta har genomgående visat sig vara fallet, något som dokumenteras i detalj i denna manual.

Tabell 6.1 Reliabilitetsvärden, medelvärden och spridningar, samt andel variansen som förklaras av skönmålning

Skala	Antal uppgifter eller ingående variabler för de aggregerade måtten	Reliabilitet (homogenitet)	Genomsnittlig korrelation mellan uppgifterna som ingår i skalan	Medelvärde	Standardavvikelse	Andel av variansen som förklaras av skönmålning
Utåtvändhet	10	0,86	0,49	3,63	0,78	0,139
Vänlighet	11	0,67	0,21	3,74	0,60	0,258
Emotionell stabilitet	8	0,75	0,34	3,54	0,75	0,276
Öppenhet	9	0,67	0,19	3,53	0,61	0,104
Noggrannhet	14	0,75	0,20	3,68	0,64	0,098
Uthållighet	8	0,78	0,23	3,80	0,65	0,112
Samarbetsvilja	11	0,79	0,18	3,91	0,71	0,243
Positiv grundhållning	10	0,84	0,28	3,93	0,69	0,338
Självförtroende	9	0,75	0,29	3,96	0,59	0,273
Social förmåga	8	0,78	0,23	3,99	0,61	0,153
Emotionell intelligens, självrapport	15	0,75	0,15	3,83	0,60	0,207
Emotionell intelligens, emotionsbedömningar	32	0,75	0,45	5,10	1,56	0,000
Kreativitet	8	0,75	0,33	3,92	0,59	0,023
Perfektionism	10	0,77	0,26	3,65	0,62	0,042
Arbetsstillfredsställelse	3	0,92		3,94	1,01	0,101
Arbetsvilja	7	0,85	0,44	3,60	0,84	0,136
Resultatorientering	13	0,70	0,11	3,91	0,49	0,058
Förändringsvilja	8	0,70	0,09	4,22	0,51	0,071
Arbetsintresse	6	0,78	0,42	3,45	0,80	0,106
Balans arbete – övrigt liv	9	0,90	0,53	3,87	0,87	0,098
Skönmålning 1 (kovert)	12	0,69	0,11	3,82	0,61	-
Skönmålning 2 (overt)	9	0,71	0,25	3,49	0,71	-
Aggregerade variabler						
Jagstyrka, okorrigerade data	11	0,90	0,44	3,80	0,45	0,300
Jagstyrka, korrigerade data	10	0,68	0,16	0,00	0,54	-
Chefspotential, korrigerade data	13	0,80	0,24			-
Stresskänslighet, korrigerade data	7	0,61	0,18			-

Den genomsnittliga reliabiliteten för de enskilda skalorna var 0,75 (median). Som framgår av Bilaga 4 har andra datauppsättningar gett mycket likartade resultat.

7. Skalornas egenskaper: stabilitet

Stabiliteten har studerats i två studier (3 och 4), i båda fallen med deltagare från psykologprogrammet vid Uppsala universitet. Testningar skedde med 6 veckors mellanrum. Resultaten för enskilda skalor ges i Tabell 7.1. Fig. 7.1 ger resultatet för jagstyrka i den ena av

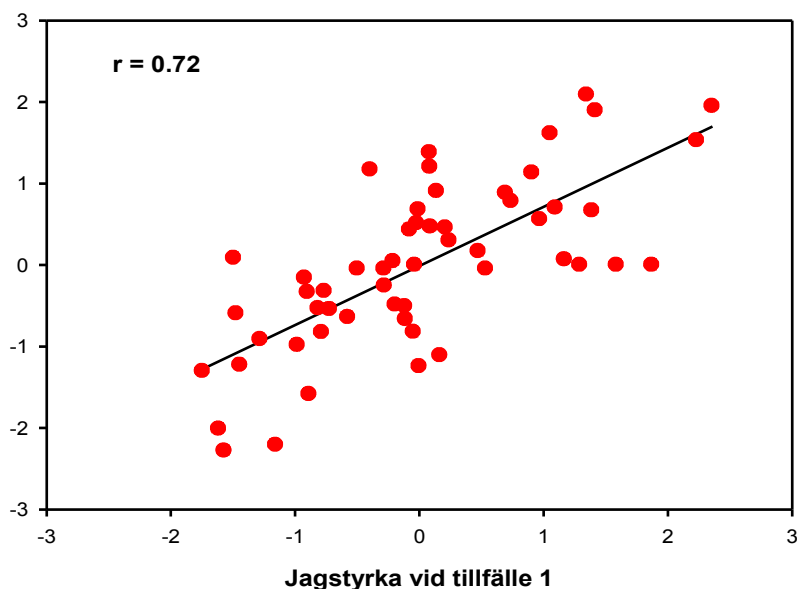
de två grupperna – resultatet i den andra gruppen var snarlikt med en korrelation på samma nivå.

Tabell 7.1 Stabilitet, Studie 3.

Variabel	Korrelation mellan tillfällena
Extraversion	0,88
Social förmåga	0,73
Emotionell stabilitet	0,60
Noggrannhet	0,81
Uthållighet	0,66
Positiv grundhållning	0,53
Emotionell intelligens	0,72
Självförtroende	0,62
Passiv aggression	0,67
Arbetsvilja	0,63
Resultatorientering	0,68
Attityd till förändringar	0,78
Kovert taktiskala	0,73
Jagstyrka	0,72

Det framgår av tabellen att testets olika variabler hade acceptabel till god stabilitet. Den genomsnittliga stabiliteten var 0,75 (median för skalorna som mäter enskilda variabler). Värdena för jagstyrka har prickats i Fig. 7.1. Eftersom jagstyrkans reliabilitet ligger på ca 0,7 tyder dessa data på en i det närmaste perfekt korrelation mellan de sanna värdena i jagstyrka vid de två tillfällena.

Jagstyrka vid tillfälle 2



Figur 7.1. Stabilitet över sex veckor, Studie 3. Jagstyrkan standardiserad till medelvärde=0 och standardavvikelse=1.

Resultaten i Studie 4 framgår av Tabell 7.2.

Tabell 7.2. Stabilitet, Studie 4.	
Testvariabel	Korrelation mellan tillfällena
Vänlighet	0,72
Öppenhet	0,82
Extraversion	0,90
Social förmåga	0,75
Emotionell stabilitet	0,72
Noggrannhet	0,83
Uthållighet	0,81
Positiv grundhållning	0,69
Emotionell intelligens	0,71
Självförtroende	0,70
Kreativ attityd	0,88
Samarbetsvilja	0,58
Perfektionism	0,87
Jagstyrka ¹	0,75
Arbetsvilja	0,64
Arbetsintresse	0,77
Resultatorientering	0,76
Attityd till förändringar	0,71
Balans	0,86
Overt taktikskala ²	0,80
Kovert taktikskala ²	0,71

Not 1. Variabel som baseras på medelvärdet av ett antal standardiserade delkomponenter.

Not 2. Dessa två variabler användes för att korrigera alla andra variabler men har inte själva korrigerats.

Den genomsnittliga stabiliteten var 0,75 (median), räknat över samtliga 22 delskalor. Jagstyrka ingick ej i detta medelvärde. Värdet är normalt för personlighetstest och för det studerade tidsintervallet [90].

De arbetsrelaterade skalorna har en stabilitet i nivå med övriga skalor. Enligt instruktionerna skulle de som tog testet betrakta sina studier som sitt arbete. De hade omfattande studieerfarenhet. Det finns ingen anledning att tro att stabiliteten hos dessa skalor skulle vara lägre i en yrkesverksam grupp.

8. Skönmålning

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 [33; 47]. Frågan är vad resultaten på ett sådant test egentligen är värda, om svaren på testet är starkt korrelerade med skönmålning, eller 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 [78]. 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 enorma.

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 [69]. Empirisk forskning visar dessutom att de ipsativa formaten inte effektivt motverkar tendensen till skönmålning [64]. I en bilaga till den tekniska rapporten ges en mera omfattande diskussion av ipsativa testformat (Bilaga 3).

Två skönmålningsskalor⁷ utvecklas och används i denna rapport. Traditionellt mäter man skönmålning, eller tendensen att ge taktiksvår, med hjälp av en skala av den typ som Crowne och Marlowe publicerade för ganska länge sedan [23], 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 [81-82]. Han använde samma typ av uppgifter som Crowne och Marlowe 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 ljuger för sig själv. Ett exempel är:

”Det är svårt för mig att stänga ute en störande tanke.”

Lönnqvist et al. [61] 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. Hans skala som byggde på frågor om normbrott korrelerade 0,84 med Crowne-Marlowe, medan skalan på ”självbedrägeri” 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. (Detta är emellertid inte så lätt att göra).

I arbetet med *UPP*-testet har konstruerats en egen skala på samma principer som Crowne-Marlowe och Paulhus, en overt skönmålningsskala (öppen mätning), men också en kovert skala (dold mätning). Den senare består av uppgifter av samma typ som i den övriga delen av

⁷ 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.

testet, men de är särskilt utvalda för att korrelerade högt med den overta taktikskalan. På så sätt får man en möjlighet att mäta taktiksvår som inte kan avslöjas av sofistikerade testtagare. De två taktikskalorna (overt och kovert) har begreppsvaliderats, se Tabell 8.1. Båda korrelerade högt med Crowne-Marloweskalen och högt sinsemellan. Den senare korrelationen var något lägre vilket är rimligt.

Tabell 8.1. Korrelationer mellan de tre skönmålningsskalorna (N=159), data från Studie 1.

	Crowne-Marlowes skala	Overt skönmålningsskal a	Kovert skönmålningsskal a
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 taktikskalan 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), 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 inte [126]. 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 taktiksvartendens 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 Fig.8.1.

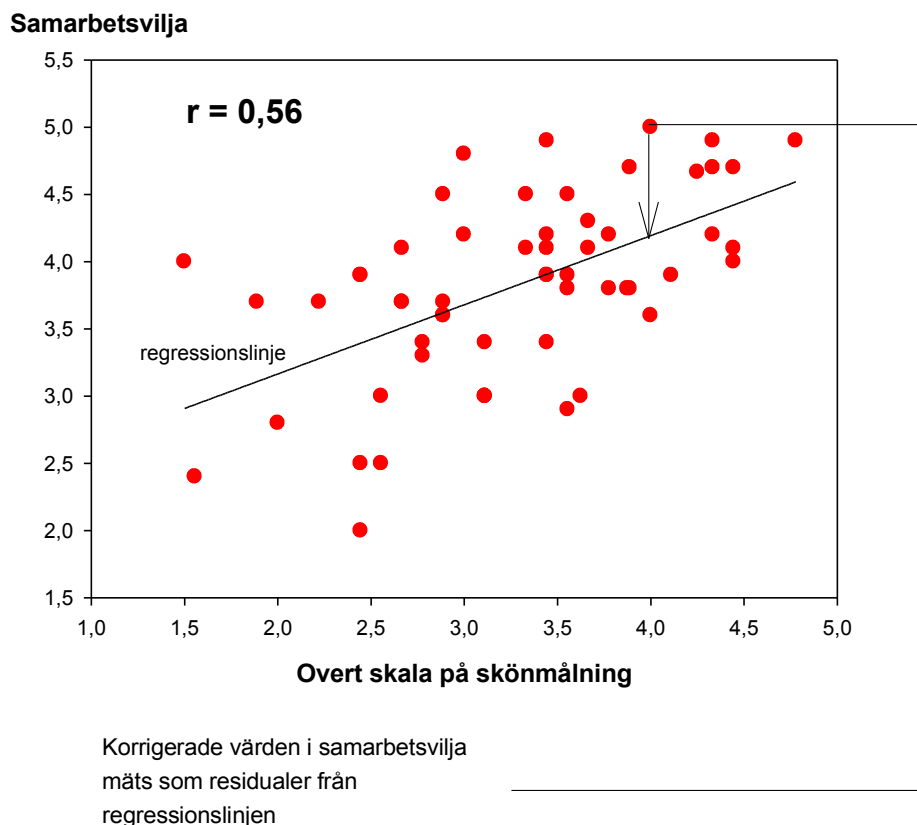


Fig. 8.1. Illustration av hur korrigerade värden i ett test uppskattas med hjälp av en psykometrisk modell.

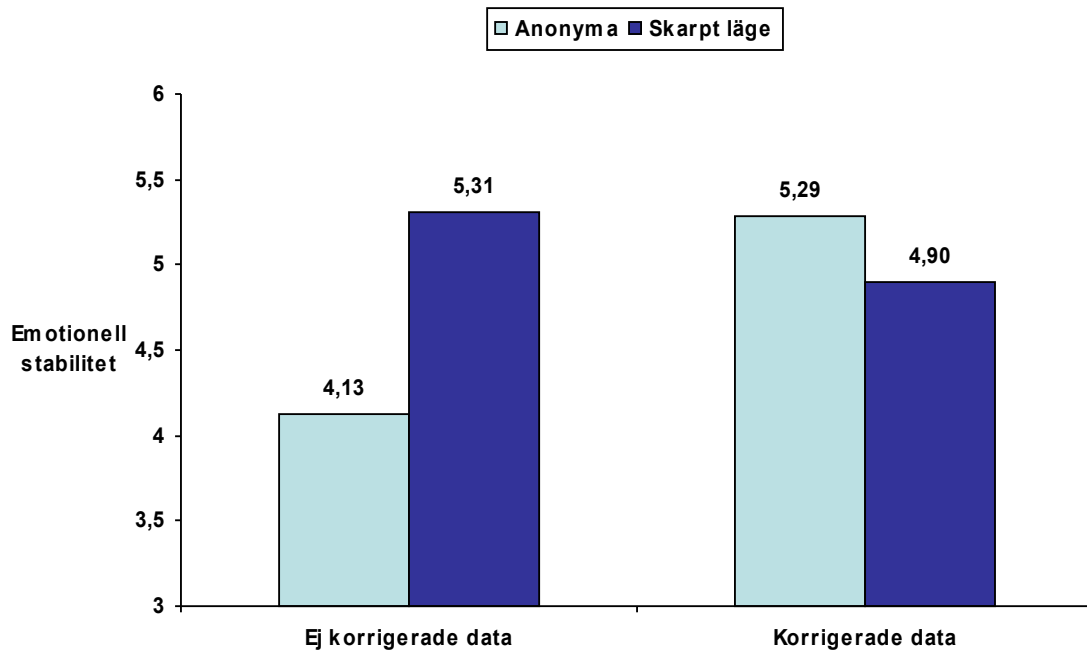
Beräkningen av residualer görs enligt vanlig regressionsanalys, dvs

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 [21] för en utförlig och avancerad diskussion.

Denna metodik utnyttjar skönmålningsskalan maximalt. Effekten av ett starkt incitament att svara taktiskt nästan helt försvinner med denna typ av korrektion [103; 112]. Skönmålningen har alltså inte längre någon effekt när korrektionen har gjorts.

Den här tekniken har fungerat utmärkt och validerats i en studie av sökande till HHS, 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 [29]. Se Figur 8.2 som visar att skillnaden i okorrigerade värden starkt reduceras när vi genomfört korrektionen för skönmålning.

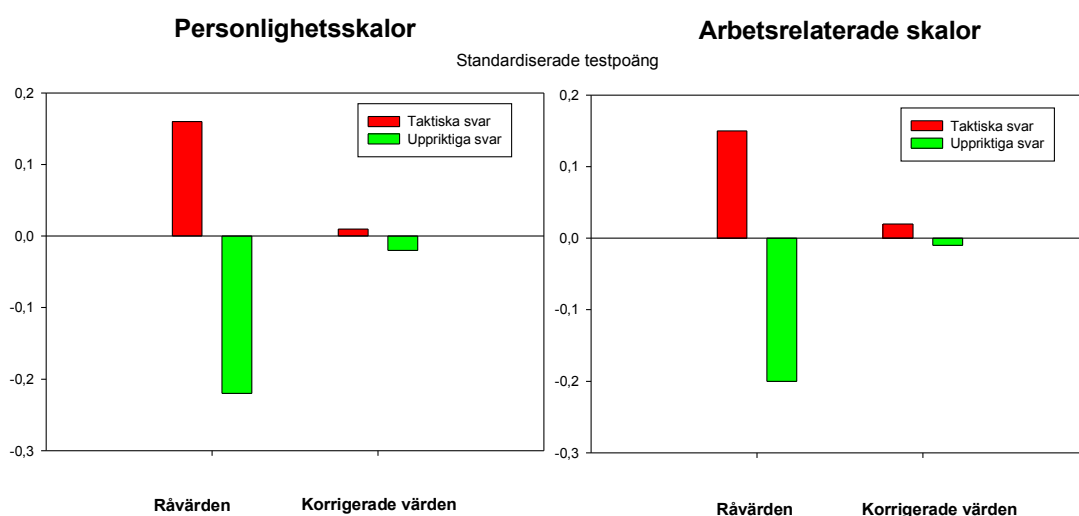


Figur 8.2. Effekten av korrektion för skönmålning vid jämförelse mellan anonym testning och skarpt läge, emotionell stabilitet. Medelvärde=5, standardavvikelsen=2.

Fig. 8.3 visar resultatet av ett experiment som beskrivs närmare i den tekniska rapporten. 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:

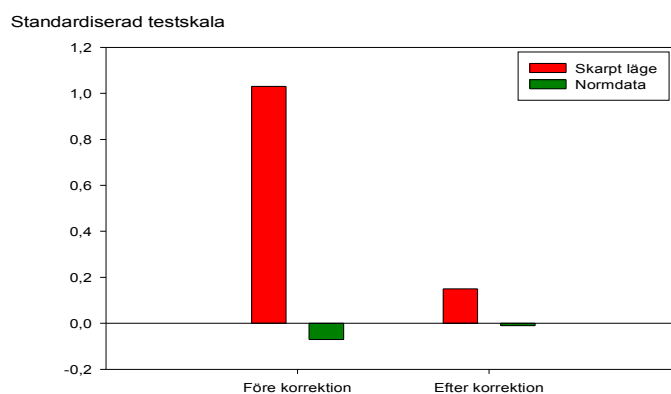
”Tänk på ett jobb som du mycket gärna skulle vilja ha. Tänk dig nu att du sökt just det jobbet och att den här testningen är ett *mycket viktigt* led i anställningsförfarandet. Svara på testuppgifterna så att framstår som en person som är precis sådan som man söker till jobbet. Det kanske innebär att 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 Fig. 8. 3.



Figur 8.3 Genomsnittliga effekter av korrektion för skönmålning, standardiserade skalor.

Delar av *UPP*-testet (screening-modulen) används i rekryteringsplattformen Katapult⁸. De som där tar testet gör det i skarpt läge, dvs. i samband med att de söker ett jobb. Se Fig. 8.4 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. Ca 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 8.4. 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 chefsrekrytering att gynnas av att man genomför en korrektion [101], 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 de alltså två skäl att korrigerar för skönmålning:

- Testvärdena ger efter korrektion en sannare bild av testpersonen eftersom eventuella försök att svara taktiskt i stort sett inte påverkar dem
- Testvärdena är mindre påverkade av slumpfel i mätningarna

UPP-testet använder alltid korrektion för skönmålning enligt den metod som beskrivs i denna manual, den tekniska rapporten och i [105].

⁸ www.katapult.com

9. Testdatas kvalitet

Kvalitetsmätning

Om man kan dra slutsatser från resultatet på ett personlighetstest beror på datas kvalitet. Det är ett stort problem vid testning att kvaliteten varierar från person till person. Somliga testade har haft alltför lågt engagemang i testningen eller missförstått uppgifterna på ett eller annat sätt. Då blir resultaten osäkra eller kanske helt oanvändbara. Det är därför viktigt att mäta datakvaliteten. I *UPP*-testet finns metodik för sådan mätning.

Det finns ett flertal aspekter på testdatas kvalitet, förutom skönmålning, att ta hänsyn till:

- Differentiering, dvs. spridning av svaren; motsatsen är att den testade personen har haft en tendens att svara med samma alternativ på alla testuppgifter
- Tendensen att instämma i testuppgifter med positivt innehåll och avvisa uppgifter med negativt innehåll
- Slumpmässiga svar, eller svar som tyder på att den testade uppfattat testuppgifterna på annat än det avsedda sättet

Skälen till att välja dessa tre mått var följande.

- Personer som har en tendens att välja samma svarsalternativ för alla frågor (liten intra-individuell spridning), t ex mittalternativet, är troligen sämre motiverade och deras testresultat bygger på mindre information
- Personer som håller med om många positiva påståenden och avvisa många negativa är troligen sämre motiverade och deras testresultat är därför osäkra. Denna variabel benämns *jasägartendens*⁹
- Personer som svarar slarvigt, slumpmässigt eller utifrån en särpräglad egen uppfattning om testfrågorna ger svar som är svagt korrelerade med gruppens genomsnittliga svar. Observera att detta mått är oberoende av *nivån* på svaren; det handlar om ett mått på avvikande struktur i svarsmonstret.

Jasägartendensen beräknades dels som det genomsnittliga svaret på alla testuppgifter som var formulerade i positiv riktning, dels som motsvarande mått på negativt formulerade testuppgifter. Dessa två mått korrelerade -0,40. Ett samlat mått på jasägartendens skapades genom att beräkna deras medelvärde, efter vändning av de positivt¹⁰ formulerade uppgifterna.

De fyra måtten på testkvalitet beräknades för alla testade personer i testdatabasen, 973 stycken. Alla variabler var korrigerade för skönmålning. Se Tabell 9.1 för sambanden mellan de fyra kvalitetsmåten. Tabellen visar också reliabiliteten för de tre kvalitetsmåten. Den har beräknats som korrelationen mellan måtten beräknade på var sin halva av mängden av

⁹ I avsnitt 18 och i den tekniska rapporten används termen *jasägartendens* för ett mått på *jasägande* i global mening, dvs. tendensen att hålla med oavsett innehåll. Denna ansats visade sig fungera mindre väl som kvalitetsmått. Vi använder termen *globalt jasägande* för att skilja den från *jasägartendens* som den definierats bland kvalitetsmåten.

¹⁰ Vändningen gjordes i denna riktning beroende på att svarsskalan kodades som 1=håller absolut med och 5=håller absolut inte med.

testuppgifter, uppjusterad enligt Spearman-Browns formel, s.k. split-half reliabilitet [63]. Stabilitetsvärdena bygger på Studie 3 och 4, sammanlagt 107 testade personer.

Tabell 9.1. Korrelationer mellan kvalitetsmått, reliabilitet och stabilitet.

	Intra-individuell korrelation	Låg ¹ ja-sägartendens	Intra-individuell spridning)	Reliabilitet	Stabilitet
Intra-individuell korrelation	1,00	0,84	0,40	0,75	0,84
Låg ¹ ja-sägartendens	0,84	1,00	0,69	0,92	0,82
Intra-individuell spridning	0,40	0,69	1,00	0,94	0,84

Not¹. Indikerar att denna variabel vändes före korrelationsberäkningarna.

Tabellen visar att de tre måtten samvarierade som väntat. De har använts för att bilda ett sammansatt mått på datakvalitet. Alfa värdet för detta mått var 0,87. En rimlig tolkning är att detta kvalitetsmått alla ger uttryck för den testade personens motivation och det engagemang och den omsorg som han eller hon investerat i uppgiften.

Kvalitetsmättet har begreppsvaliderats. Det samvarierade med:

- Ålder – äldre testade hade högre datakvalitet
- Kön – kvinnor hade högre datakvalitet
- Utbildning – högre utbildning var förknippad med högre datakvalitet
- Testningens konsekvenser – datakvaliteten var högre vid testning i skarpt läge

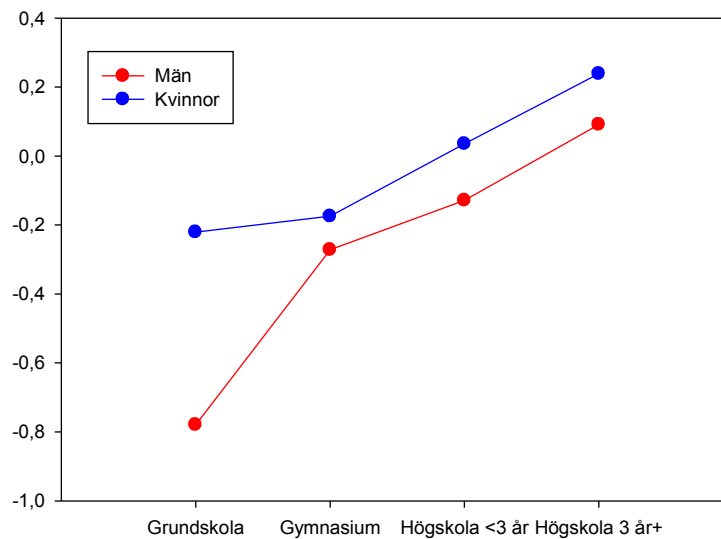
Dessa samband är rimliga och är begreppsvalideringar av mätningen av datakvalitet. Samtliga var statistiskt signifikanta. Spannet i datakvalitet (standardiserade värden) över variationen i de oberoende variablerna var:

- Ålder – 0,30 (5 åldersgrupper)
- Kön – 0,12
- Utbildning – 0,35 (4 utbildningsgrupper)
- Testningens konsekvenser – 0,68

Den absolut största betydelsen hade alltså testningens konsekvenser. Om dessa var viktiga lade de testade ner större engagemang i sina svar, vilket ju är rimligt.

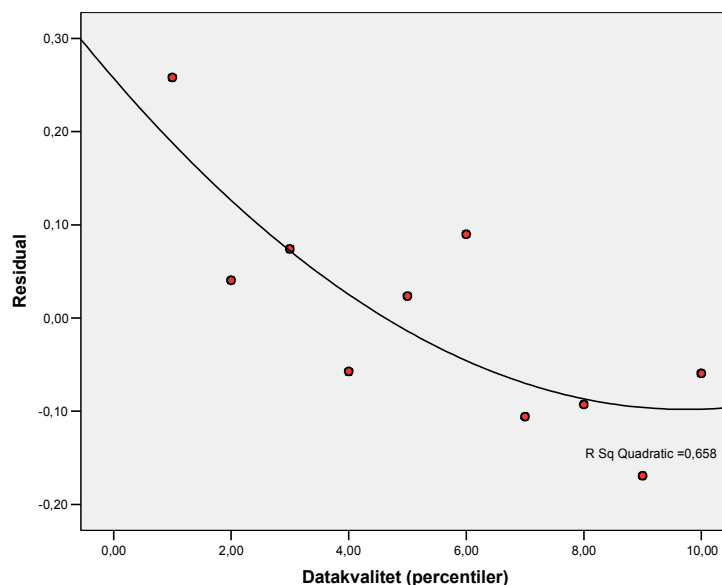
Figur 9.1 visar relation mellan datakvalitet, kön och utbildning.

Datakvalitet



Figur 9.1. Datakvalitet, kön och utbildningsnivå.

För att studera datakvalitetens betydelse bildades ett sammansatt mått av proxykriterierna (arbetsmotivation, arbetsintresse, arbetstillfredsställelse, resultatorientering och förändringsvilja men inte balans). Den multipla korrelationen för prognos av detta sammansatta kriterium var 0.60 med samtliga personlighetsvariabler som oberoende variabler. Residualen i regressionsekvationen beräknades för varje person. Den genomsnittliga residualen beräknades sedan för percentilgrupper i datakvalitet, se Fig. 9.2. Figuren visar det väntade sambandet: lägre datakvalitet gick samman med större residual, alltså sämre förmåga att förklara eller predicera kriterierna. Detta är ytterligare stöd för vår metodik att mäta datakvalitet.



Figur 9.2. Residual i regression av sammansatt kriterium mot samtliga personlighetsvariabler. Residualen i absolutbelopp och efter standardisering.

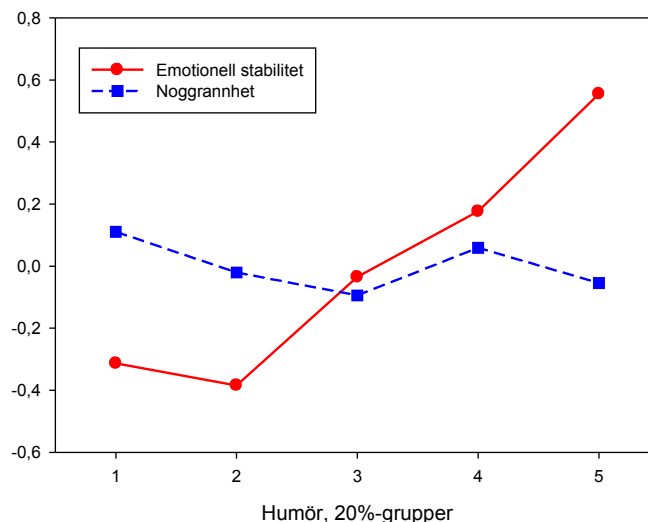
Vid den praktiska användningen av testet integreras information om testdatas kvalitet i testrapporten för varje person, tillsammans med de två måtten på skönmålning och information om antalet ej besvarade testuppgifter. Om värdet på kvaliteten är lågt, 2 eller mindre på den 9-gradiga stanineskalan, finns anledning att behandla testresultaten mycket försiktigt, kanske att göra om testningen efter utredning av skälen till den låga kvaliteten. Fig. 9.2 tyder på att det är särskilt de mycket låga kvalitetsvärdena som bör ge anledning till försiktighet.

Stämningläge

Psykologer har av tradition betraktat personligheten som tämligen konstant, och så även personligheten som den mäts med test. Men påverkas inte resultaten av det emotionella tillstånd som den testade befinner sig i?

Fig. 9.3 visar ett exempel från testning i mycket skarpt läge. Den gäller 190 sökande till Handelshögskolan i Stockholm som genomgick ett omfattande testprogram som avsåg att mäta personlighet i en mängd olika avseenden. Det är rimligt att anta att det för många av dem var en mycket viktig testsituation. Inte bara vanliga testdata togs in, utan också bedömningar av stämningläge (humör), se [116] för en förklaring av metodiken för att mäta stämningläge, som fö har kommit till mycket stor användning och begreppsvaliderats, se t ex [96]. Figuren visar, som ett exempel bland många möjliga, på sambandet mellan stämningläge vid testningens början och resultat: i detta fall hedonisk ton (glad - ledsen) och utåtvändhet. Som synes finns ett tydligt samband för emotionell stabilitet men inget för noggrannhet.

Medelvärde i personlighetsskala, standardiserad



Figur 9.3. Emotionell stabilitet och noggrannhet som funktioner av stämningläge vid testningen.

Om man ska tolka ett testresultat är det tydligen mycket viktigt att veta hur den testade kände sig vid testningen. Ett tillfälligt känslotillstånd kan ha medfört att hans eller hennes resultat på testet blivit starkt missvisande. Detta gäller i synnerhet personlighetstest, och "skarpa lägen",

dvs. när mycket står på spel. Figuren visar att stämningläge kan vara av stor betydelse för vissa testvariabler men inte för andra.

Tabellerna 9.2 och 9.3 ger mera detaljer beträffande sambandet mellan personlighetsvariabler och stämningläge vid testningen.

Tabell 9.2. Korrelationer mellan stämningläge och Big Five-dimensionerna, ej korrigerade för skönmålning.

	Glad - ledsen	Pigg - trött	Lugn - spänd	Humör, sammanslaget
Vänlighet	0,09	0,11	0,08	0,11
Emotionell stabilitet	0,50**	0,35**	0,38**	0,50**
Öppenhet	0,38**	0,50**	0,21**	0,41**
Utåtvändhet	0,43**	0,37**	0,22**	0,39**
Noggrannhet	0,01	0,30**	0,03	0,11

** $p < 0.01$

Tabell 9.3. Korrelationer mellan stämningläge och Big Five-dimensionerna. Alla värden korrigerade för skönmålning.

	Glad – ledsen	Pigg – trött	Lugn – spänd	Humör, sammanslaget
Vänlighet	0,10	0,10	0,08	0,12
Emotionell stabilitet	0,38**	0,15*	0,19**	0,31**
Öppenhet	0,24**	0,39**	0,01	0,28**
Utåtvändhet	0,35**	0,27**	0,09	0,31**
Noggrannhet	-0,16*	0,17*	-0,14*	-0,06

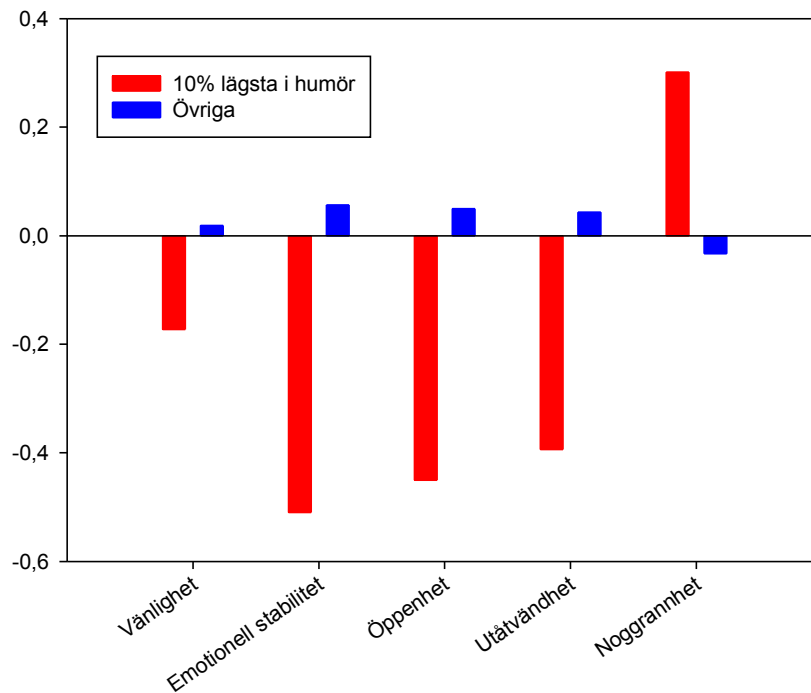
* $p < 0.05$ ** $p < 0.01$

Båda tabellerna visar på starka eller ganska starka samband mellan vissa personlighetsvariabler och stämningläge, kanske främst hedonisk ton och vakenhet ("activation"). Det är intressant att se att stämningläge tycks ha varit av mindre betydelse för testvariabler korrigerade för skönmålning. Test som inte använder korrektion kan därför förväntas vara mera utsatta för denna felkälla.

Sambanden mellan skalvärden och humör vid testningens början var tydliga i åtminstone tre av de fem personlighetsdimensionerna. Den dominerande trenden var att ett lägre värde i stämningläge åtföljdes av ett lägre värde i personlighetsdimensionerna.

Fig. 9. 4 visar medelvärden i FFM-skalorna för de 10 % med "sämst" humör och för övriga.

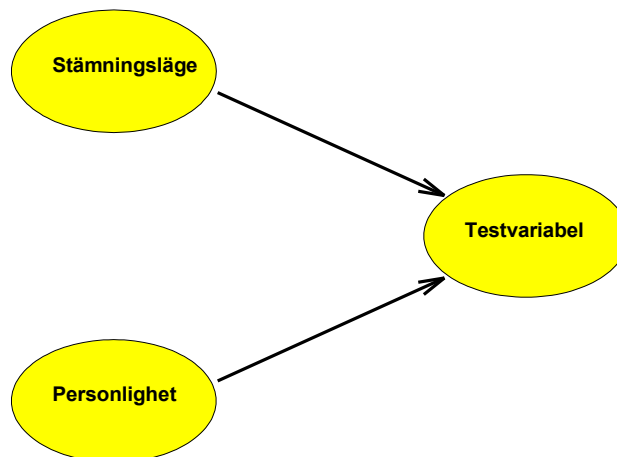
Medelvärde i standardiserade personlighetsskalor



Figur 9.4. Skillnader i FFM-skalor mellan de 10 % med ”sämst” humör och övriga.

Det är intressant att se att värdena var låga för dem som var emotionellt ur slag, för fyra av de fem FFM-skalorna men höga för den femte, nämligen noggrannhet.

Dessa resultat kan tolkas på olika sätt. Ett förslag ges i Figur 9.5



Figur 9.5. Modell för sambandet mellan stämningläge, personlighet och testvariabler.

Andra möjligheter finns givetvis, t ex att stämningläge medierar sambandet mellan personlighet och testvariabeln. Oavsett vilken modell man ansluter sig till blir den praktiska slutsatsen att testning som genomförts i ett extremt stämningläge kan ge missvisande resultat. Därför är det viktigt att testningen innehåller frågor om stämningläge. Svaren på sådana frågor kan vara vägledande för att värdera denna viktiga aspekt på datakvalitet.

UPP-testet börjar med 12 frågor som mäter stämningsläge: 4 frågor vardera för dimensionerna glad-ledsen, pigg-trött och lugn-spänd. Dessa utvärderas med hjälp av en särskild normdatabas för stämningsläge och sammanvägs till ett mått på humör. Reliabiliteterna (alfa) för de tre skalorna på stämningsläge samt det aggregerade måttet humör var:

- Glad-ledsen: 0,73
- Pigg-trött: 0,80
- Lugn-spänd: 0,79
- Humör (aggregerat mått på stämningsläge): 0.60

Attityd till testet: "Face validity"

Attityd till testet och testningen är av betydelse vid tolkningen av testresultaten och kan behöva tas upp i återkopplingen. Därför har ett antal frågor som tar upp attitydaspekter lagts till sist i testet. Se tabellerna 9.4 och 9.5

Tabell 9.4. Bedömning av UPP-testet av de testade, procent.

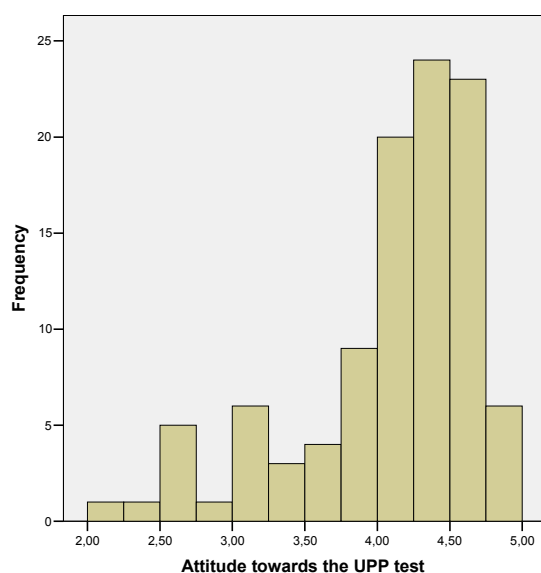
	Nej, absolut inte	Nej, i stort sett inte	Tveksamt	Ja, i stort sett	Ja, absolut
Anser du att testningen var meningsfull?	8,3	8,3	6,9	44,4	31,9
Tog testet upp viktiga frågor?	8,3	6,9	1,4	56,9	26,4
Var instruktionerna klart formulerade?	1,4	2,9	1,4	24,3	70,0
Var testuppgifterna klart formulerade?	2,8	1,4	4,2	40,3	51,4
Var svarsalternativen klart formulerade?	2,8	2,8	5,6	45,1	43,7
Var det en intressant uppgift att besvara testuppgifterna?	7,0	7,0	2,8	29,6	53,5
Var det något som saknades i testet?	21,1	32,4	42,3	4,2	0,0

Tabell 9.5. Jämförelse mellan UPP-testet och andra test, procent.

	Har inte tagit något annat test	Klart bättre	Något bättre	Varken bättre eller sämre	Något sämre	Klart sämre
Var testet bättre eller sämre än andra personlighetstest	11,3	12,7	26,8	40,8	8,5	0,0

Tabellerna visar att UPP-testet bedömdes positivt av dem som tog testet. Det hindrar inte att negativa reaktioner fanns i vissa fall. Det kan vara viktigt att beakta vid tolkningen testresultaten. Ett sammanslaget mått på attityd till testet ingår därför i rapporten över testningen. Det är medelvärde av de 8 attitydfrågorna, se Tabellerna 8.4 och 8.5. Den sista frågan, Tabell 9.5, konverterades först genom att betrakta svaret ”har inte tagit något annat test” som uteblivet svar (”missing”), varefter frågan vändes så att riktningen av positiva-negativa svar skulle bli densamma som för övriga frågor. Reliabiliteten (Cronbachs alfa) för skalan på attityd till testet och testningen var 0.86.

Fördelningen av den genomsnittliga attityden till *UPP*-testet bland dem som tagit testet och gjort sådana bedömningar (t o m mars 2010)¹¹ framgår av Fig. 9.6. Figuren visar på övervägande mycket positiva bedömningar av testet.



Figur 9.6. Fördelning av attityder till *UPP*-testet bland dem som tagit testet.

10. Begreppsvalideringar

Här redogörs för begreppsvalideringar som genomförts genom att korrelera testvariabler med andra variabler som mäter samma eller snarlika begrepp. De senare har lokaliserats i den internationella forskningen. Det handlar i de flesta fall inte om andra test utan om mätprocedurer som publicerats internationellt och är allmänt tillgängliga. I Bilaga 2 redogörs för ytterligare begreppsvalideringar. Bilaga 3 ger en teknisk diskussion av statistiska problem vid valideringar.

Vid begreppsvalidering vill man veta om en testvariabel verkligen mäter det begrepp den utger sig för att mäta. Det är därför nödvändigt att korrigera korrelationerna mellan testvariabeln och kriterierna för mätfel. Se vidare Bilaga 3. Mätfelen har den effekten att de sänker alla korrelationer. I det följande rapporteras både korrelationer mellan rådata och

¹¹ Frågorna om attityd till testet infördes hösten 2009.

korrelationer efter korrektion för slumpmässiga mätfel. Som påpekas i Bilaga 3 är det ibland lämpligt att göra andra typer av analyser än korrelationsberäkningar vid begreppsvalidering. Så har också skett vad gäller *UPP*-testet. I dessa fall kan en sammanställning av resultaten som enbart bygger på korrelationsdata vara missvisande.

Personlighetsvariabler

I testet ingår de 5 personlighetsvariablerna i FFM samt 8 mera specifika variabler orienterade mot arbetsfunktioner. Dessutom estimeras chefslämplighet, stresskänslighet och jagstyrka, alltså sammanlagt 16 personlighetsvariabler. Samtliga dessa 16 variabler har begreppsvaliderats, antingen genom korrelationsanalyser och eller genom andra typer av analyser. Resultaten av utförda begreppsvalideringar summeras i Tabell 10.1.

Tabell 10.1. Sammanställning av begreppsvalideringar av personlighetsvariabler			
Variabel	Begreppsvaliderad mot	Korrelation med begreppskriteriet	Korrelationer efter korrektion för mätfel i båda de ingående variablerna
Utåtvändhet	Utåtvändhet IPIP	0,76 (Studie 2), 0,79 (Studie 3)	0,76 → 0,88 0,79 → 0,93
Emotionell stabilitet	Emotionell stabilitet IPIP (Studie 2 och 3)	0,69 (Studie 2), 0,70 (Studie 3)	0,69 → 0,86 0,70 → 0,85
Öppenhet	Öppenhet IPIP	0,72 (Studie 4)	0,72
Vänlighet	Vänlighet IPIP	0,57 (Studie 4)	0,57 → 0,78
Noggrannhet	Noggrannhet IPIP (Studie 2 och 3)	0,66 (Studie 2), 0,71 (Studie 3),	0,66 → 0,92 0,71 → 0,93
Samarbetsvilja	Hogan skala för passiv aggression (Studie 3),	-0,42 (Studie 3),	-0,42 → -0,66
Kreativitet	Öppenhet IPIP	0,68 (Studie 4)	0,68
Emotionell intelligens	Schuttes et al. EI-skala (Studie 3), Furnhams TEIQue [83] (Studie 15)	0,36 (Studie 3), 0,54 (Studie 15)	0,36 → 0,50 0,54 → 0,68
Social förmåga	UCLA:s ensamhetsskala[89] (Studie 5+6) Jones et al.: shyness-skala, egen skala för social osäkerhet (PM-SOC) (Studie 15)	(Studie 5+6, ensamhet) 0,43 (Syness, Studie 15) 0,46 (PM-SOC Studie 15)	0,58 → 0,74 0,52 → 0,80 0,43 → 0,59 0,46 → 0,68
Grad av positiv grundattityd	Affektskalor av Tsai, Knutson och Fung [120] (Studie 5+6)	0,46 (Studie 5+6, positiva emotioner), -0,32 (Studie	0,46 → 0,55 -0,32 → -0,39

Tabell 10.1. Sammanställning av begreppsvalideringar av personlighetsvariabler

		5+6, negativa emotioner)	
Uthållighet	Gritskalan [25](Studie 11)	0,37	0,37 → 0,51
Självförtroende	Rosenberg [88]	0,57 (Studie 2), 0,51 (Studie 3)	0,57 → 0,70 0,51 → 0,66
Perfektionism	CMD-skalan [32] (Studie 10)	0,30	0,30 → 0,36
Jagstyrka	Härdighet (hardiness) [62], samt ett mått på föregripande inställning (proactivity) [5] ¹²	0,65 med härdighet (Studie 4) och 0,52 med föregripande attityd (Studie 4)	0,65 → 0,77 0,52 → 0,62
Chefpotential	Chefskarriär [Studie 13 [106; 115]], <i>d</i> -värde	0,52	-
Stresskänslighet	Begreppsvaliderad mot relationsproblem vid skilsmässa, Studie 13 och Bilaga 2, <i>etavärde</i>	0,49	-

Det genomsnittliga värdet (medianer) för de okorrigerade korrelationerna (absolutbelopp) var 0,54; för de korrigerade 0,68¹³. Eftersom dessa valideringar syftar till att studera sambandet mellan de underliggande begreppen bakom skalorna, inte skalorna själva, är värdena i den fjärde kolumnen i Tabell 10.1 de relevanta.

Den enda skala som kan ifrågasättas på grundval av begreppsvalideringarna är perfektionismskalan. CMD-skalan, det använda kriteriet för perfektionism, är emellertid ett ganska indirekt mått på perfektionism som tycks utgå från ett antagande om att motivet till perfektionism är att man är rädd för ”vad andra ska tänka” om man misslyckas eller visar sig vara slarvig. Det är intressant att notera att CMD var starkt laddat med skönmålning. Korrelationerna mellan CMD och skönmålning var 0,65 och 0,52 för kovert och overt skönmålning. Motsvarande värden för perfektionismskalan i testet var 0,14 och -0,09. I förhållande till noggrannhet korrelerade CMD-skalan 0,25 medan perfektionismskalan hade ett starkare samband, nämligen 0,36. Båda måtten korrelerade som väntat negativt med kreativitet: -0,34 för CMD och -0,31 för perfektionismskalan.

Det finns alltså ett visst stöd för att perfektionismskalan i testet är en bättre skala för mätning av tvångsmässig noggrannhet än CMD-skalan. Ett lågt samband med skönmålning är önskvärt och CMD-skalans starka samband med skönmålning är onekligen ett memento, liksom dess ganska låga samband med noggrannhet. Perfektionism-skalan i *UPP*-testet är dessutom mera direkt och tydlig mätning av just perfektionsim, inte av det indirekta slaget som CMD-skalan är ett exempel på.

¹² Dessa två variabler var främst avsedda att begreppsvalidera jagstyrkevariabeln; övriga samband återges inte här med finns i Tabell 7. 4 i den tekniska rapporten. [106]

¹³ *d*-värdet för chefpotential ej medräknat här eftersom det inte är en korrelation, dock tyder värdet på en medelstor till stor effekt [20].

Proxy-kriterier (arbetsmotivation, balans och andra arbetsrelaterade variabler)

De fem ”egna” kriterier som kommit till användning och som testet mäter framgår av Tabell 10.2. Tabellen ger också information om begreppsvalideringar. Arbetstillfredsställelse mäts dessutom, med tre frågor som ingår i Arbetslivsinstitutets standardmätt på detta begrepp och har därför inte ansetts behöva valideras i detta sammanhang, utan har i stället använts som matchande variabel till arbetsvilja¹⁴. Alla de övriga kriterierna har konstruerats för UPP-testet.

Tabell 10.2. Sammanställning av begreppsvalideringar av arbetsrelaterade variabler (kriterier)

Variabel	Begreppsvaliderat mot	Korrelation med begreppskriteriet	Korrelationer efter korrektion för mätfel i båda de ingående variablerna
Förändringsvilja	Oregs skala för mätning av förändringsmotstånd ¹⁵ [79-80] (Studie 9)	0,59 (Studie 9)	0,59 → 0,82
Arbetsvilja	ALI:s skala för mätning av arbetstillfredsställelse	0,66 (Studie 5+6)	0,66 → 0,73
Resultatorientering	Rays skala för mätning av prestationsmotivation [84-85]	0,56 (Studie 5+6)	0,56 → 0,75
Arbetsintresse	BPS-skalan [125]	0,45 (studie 9)	0,45 → 0,55
Balans arbete – övrigt liv	Haymans balansskala [38]	0,68 (Studie 9)	0,68 → 0,76

Det genomsnittliga värdet för de okorrigerade korrelationerna (aritmetiska medelvärden, endast 5 värden) var 0,58; för de korrigerade 0,72. Eftersom dessa valideringar syftar till att studera sambandet mellan de underliggande begreppen bakom skalorna, inte skalorna själva, är värdena i den fjärde kolumnen i Tabell 10.3 de relevanta. Samtliga skalor tycks ha fungerat väl enligt dessa valideringar. Information om kvalitetsmåttens begreppsvaliditet finns i avsnitt 9.

¹⁴ Arbetsvilja och arbetstillfredsställelse är relaterade begrepp men inte synonymer; arbetsvilja torde vara mera relevant om man vill förstå och göra en prognos av arbetsresultat [7-8; 113].

¹⁵ Ersätter tidigare begreppsvalidering mot Herscovitchs skala; Oregs skala är direkt relevant vilket skalan av Herscovitch inte var.

11. Validering mot externa¹⁶ kriterier och proxykriterier

Chefsansvar¹⁷

Framgång i chefsarbete kan studeras dels när det gäller karriären, dvs. om man befordras till chef eller inte, dels prestation i chefsuppgifterna. Exempel ges av många studier av sambandet mellan chefsfunktioner och personlighet, se t ex en omfattande metaanalys [53]. Judge et al. skiljer mellan ”leader emergence” och ”leader effectiveness”. I föreliggande studie analyseras den förstnämnda aspekten. Den beroende variabeln i analyserna var binär: om man har chefsuppgifter eller inte.

Ett erbjudande om att delta i undersökningen skickades till en grupp personer som deltagit i en tidigare undersökning av riskattityder. De utlovades två trisslotter som belöning samt en rapport över sitt testresultat. De svarande tog *UPP*-testet i sin helhet. Dessutom besvarade de ett antal frågor som avsåg att utöka mätningen av emotionell intelligens och social anpassning i arbetet samt frågor om chefsansvar. Deltagarna var från början ett slumpmässigt urval av befolkningen i åldrarna 18-75 år. Av dessa hade ca 50 % sagt att de gärna ville vara med i nya undersökningar.

De som genomförde hela testningen på vår Internetplattform var till antalet 107, av 250 som fick erbjudandet. Av dessa var 55 män och 51 kvinnor¹⁸. Medianådern var 49 år. Enbart grundskola hade 12 %, gymnasieutbildning 36 %, påbörjad högskola 24 % och examen från högskola 27 %. Andelen heltidsarbetande var 69 %, genomsnittlig månadsinkomst var för hela gruppen 27 000 kronor. Fem besvarade ej frågan om de hade chefsuppgifter varför analyserna i denna rapport baseras på 102 svarande. Av dem som besvarade frågan var hade 21 chefsuppgifter.

Gruppen av svarande är inte representativ för landet eftersom den innehöll en alltför stor andel med hög utbildning. Detta är ett vanligt problem i alla survey-undersökningar. Bortsett från utbildningsnivå var gruppen sammansatt av personer med varierande ålder och en genomsnittlig inkomst på normalnivå. Data insamlades under våren och sommaren 2009.

Analysmetod

Eftersom den beroende variabeln var binär (chef/ej chef) var det lämpligt att genomföra analysen som binär logistisk regression. Denna metodik finns väl beskriven och förklarad i [21] och är tillgänglig i SPSS. I föreliggande studie användes SPSS version 13.0.

Stickprovets storlek var $n=102$. Antalet oberoende variabler var potentiellt $5+14$, alltså 5 för FFM-modellen samt 14 för de i *UPP*-testet tillkommande variablerna. Stickprovet är för litet för analys mot 19 oberoende variabler. Eftersom det var intressant att analysera dels hela testet, dels FFM-modellen och de tillkommande *UPP*-variablerna för sig, bildades två index, med lika vägning¹⁹ för ingående variabler:

¹⁶ Vad som menas med ”externa” är oklart. I EFPA:s anvisningar finner man formuleringen ”not part of the instrument”. Se Bilaga 3.

¹⁷ Ett tack till Per Näsman, KTH, som granskat den statistiska metodiken i detta avsnitt.

¹⁸ En person besvarade inte frågan om kön.

- FFM-variablerna
- *UPP*-testets tillkommande personlighetsvariabler samt *UPP*-testets mått på resultatorientering m fl arbetsrelaterade variabler

Resultat

Binära regressionsanalyser gjordes dels för hela testet, dels separat för FFM-modellen och dels för *UPP*-testets tillkommande variabler. Resultaten framgår av Tabell 11.1.

Tabell 11.1. Korrelationer mellan chefsansvar och *UPP*-testet.

Variabelgrupp	<i>R</i> , Cox & Snell	<i>R</i> , Nagelkerke	χ^2	df	<i>p</i>
Hela <i>UPP</i> -testet	0.412	0.516	18.968	3	<0.0005
FFM-modellen	0.045	0.055	0.172	1	0.678
<i>UPP</i> -testets tillkommande variabler	0.410	0.507	18.280	2	<0.0005

SPSS ger korrelationerna i kvadrat, för Tabell 11.1 har kvadratroten ur dessa kvadrerade korrelationer beräknats för att få en skala jämförbar med den vanliga korrelationsskalan.

Måttet på modellens förklaringsnivå beräknad enligt Cox & Snell är en underskattning beroende på att detta mått ej kan nå värdet 1 om inte den beroende variabeln är fördelad med exakt lika stor andel för de två värdena som variabeln kan anta, alltså $p = 0.5$. I det här aktuella fallet låg *p*-värdet i denna mening långt från 0.5. Nagelkerkes mått på förklaringsnivå korrigerar för denna faktor och är därför mera rättvisande.

Tabell 11.1 visar att hela *UPP*-testets förklaringskraft vis à vis kriteriet chefsansvar motsvarade en korrelation på 0.516. Värdet som ges av koefficienten enligt Cox och Snell, 0.412 för hela testet, är en underskattning eftersom *p*-värdet för den beroende variabeln starkt avvek från 0.5. Det kan också noteras att modellens anpassning var mycket starkt signifikant. Detta är visserligen ingen indikation på sambandets styrka, som mäts av korrelationskoefficienterna, men ändå av intresse att påpeka eftersom den studerade gruppen var av måttlig storlek.

Tabell 11.1 visar dessutom tydligt att FFM-modellen var ineffektiv för att förklara det studerade kriteriet medan de i *UPP*-testet tillkommande variablerna däremot var effektiva som förklarande variabler, med hög korrelation och hög signifikansnivå. En slutlig jämförelse har gjorts med hjälp av en hierarkisk binär logistisk regression, där FFM-variablerna fick bilda block 1 och övriga variabler i *UPP*-testet block 2. Se resultat i Tabell 11.2.

¹⁹ Perfektionism ingick med negativ vikt, liksom balans arbete/övrigt liv. Den sistnämnda variabeln mäter problem i livet som är särskilt utpräglade för chefer. I denna studie var skillnaden mellan chefer och icke chefer -0.59, i termer av standardiserade värden, korrigerade för skönmålning. En betydande skillnad med andra ord.

Tabell 11.2. Hierarkisk binär logistisk regression av kriteriet chef/ej chef mot *UPP*-testet i två block.

Steg	-2 log likelihood
Block 1	103.552
Block 1 + block 2	84.756

Måtten i kolumn 2 av tabellen är uppskattningar av modellernas brist på anpassning. Förbättringen tack vare införande av block 2 är skillnaden mellan dessa två värden, alltså 18.796. Detta värde kan tolkas som χ^2 med $df=1$, $p < 0.005$.

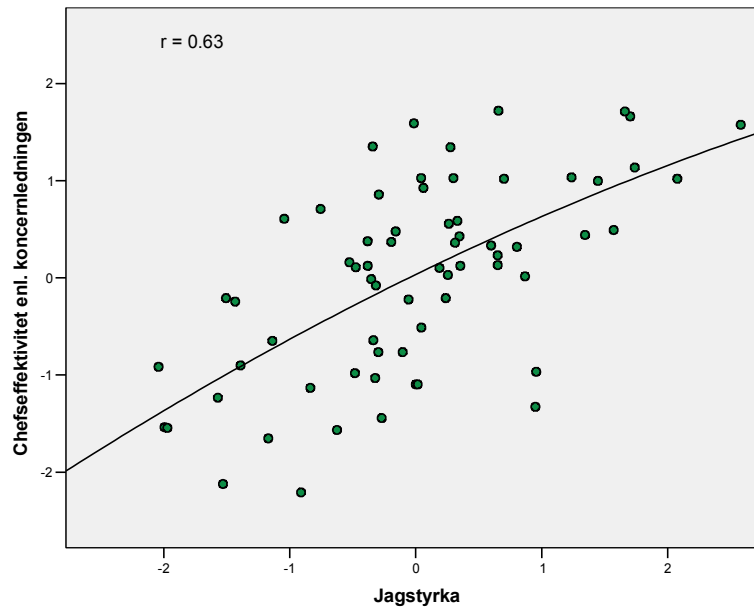
Den typ av analys som gjorts här är annorlunda än den vanliga linjära regressionsanalysen som brukar användas för att studera testdatas validitet. Bakom valet av metodik låg det förhållandet att den beroende variabeln var binär. De värden som erhållits är inte fullt ut jämförbara med dem som erhålls i vanlig linjär regression. Cohen et al. [21] diskuterar detta problem och konkluderar att måtten på förklaringskraft som erhålls med binär logistisk regression troligen är underskattningar jämförda med ordinär regression.

Analyserna har visat att:

- *UPP*-testets kriterievaliditet med avseende på chefsansvar i denna studie låg på 0.52, vilket troligen är en underskattning. Värdet är på den nivå som framkom i meta-analysen av Judge et al. var jämförbart.
- FFM-modellen saknade förklaringsvärde för det studerade kriteriet
- Tillskottet av förklaringsvärde genom *UPP*-testets variabler utöver FFM-modellen var starkt statistiskt signifikant

Chefers arbetsresultat

I en studie undersöktes relationen mellan jagstyrka hos 66 chefer (varuhuschefer) och koncernledningens bedömning av deras arbetsresultat, se Fig. 11.1. Korrelationen var hög, $r = 0.63$, $p < 0.0005$.



Figur 11.1 Samband mellan arbetsresultat och jagstyrka, 66 chefer.

Detta resultat är alltså på samma nivå, t o m bättre, som i studien av chefskarriär.

Kundservice

Inledning och syfte

UPP-testets validitet har dokumenterats tidigare, med hjälp av proxykriterier [103] och externa kriterier (chefs karriär) [106]. Syftet med denna studie var att undersöka *UPP*-testets validitet för kundservicearbete med hjälp av chefs- och förmansbedömningar samt genom fortsatta studier av proxykriterierna. Servicekvalitet är av stor betydelse för denna typ av företag [24; 35], liksom f ö i många andra sammanhang. Personlighet och attityder har tidigare befunnits vara av vikt i kundservicearbete [6; 14].

Kriterierelaterad validitet har många aspekter [57]. En viktig fråga är vilken eller vilka dimensioner som testet ska relateras till. Motowidlo, Hogan m fl. skiljer mellan kärnuppgifter i jobbet och social fungerande i arbetet [9; 74].

Normalt ses kriterierna som mått på vilken nytta den anställda gör för företaget. Kriterier kan emellertid ses också utifrån den anställdes synpunkt: lyckas han eller hon behålla jobbet och göra karriär?

UPP-testet mäter 13 personlighetsvariabler och 6 attityder till jobbet och arbetsmotivation [109]. Alla 19 dimensionerna kan inte antas korrelera med alla kriterieaspekter; det vore i själva verket förvånande om så skulle vara fallet. Vilka samband som kan förväntas beror också på vad som krävs i jobbet [43]. Om det inte ställs krav på kreativitet är det inte rimligt att vänta sig att just den variabeln ska ha hög validitet i sammanhanget.

Om man har ett stort stickprov kan man också undersöka multipla korrelationer mellan testet och ett eller flera kriterier, men det är inte fallet i den studie som rapporteras här. Det är

tidsödande och dyrbart att genomföra så stora valideringsstudier att den typen av analys kan genomföras.

Metod

Det deltagande företaget är ett finansbolag med fokus på konsumentkrediter. Anställda handläggare inom avdelningen för Customer Service inbjöds att delta i undersökningen, 65 totalt. De erbjöds arvode om 250 kr, samt om de så önskade testrapport, intyg och/eller diplom. Dessutom utlovades ytterligare 250 kr om 90 % av de inbjudna deltog. Så blev inte fallet, 53 av 65 deltog, alltså 82 %. Av dessa var 14 män och 39 kvinnor, ålder 17-29 år, genomsnitt 21.9 år. Utbildning: 33 gymnasium, 11 påbörjad högskola, 9 examen från högskola. De arbetade som kundtjänstmedarbetare, bedrägerihandläggare och inkassohandläggare.

Chefer och förmän bedömde deltagarnas arbetsinsatser och sociala funktion på jobbet enligt ett omfattande formulär som konstruerats för denna studie, se Bilaga 1. Formuläret omfattade 40 bedömningsvariabler som dels gällde övergripande kvalitativa och kvantitativa aspekter, dels mera specifika aspekter. Det är inspirerat av Hogans begrepp ”getting along and getting ahead” [43; 45].

Resultat

I ett första avsnitt redogörs för testets egenskaper i denna grupp och gruppens medelvärden, sedan följer ett avsnitt om strukturen hos kriterievariablerna och därefter beskrivs utfallet av valideringsanalyserna.

Testets egenskaper och gruppens medelvärden

Reliabiliteterna hos testvariablerna återges i Bilaga 2, rapport 2010:6. De var genomgående höga och på samma nivå som i andra studier. Värdena på kvalitetsindex framgår av Tabell 11.3 där de jämförs med normgruppen.

Tabell 11.3. Skillnader i kvalitetsmått (standardiserade värden, M=0, SD=1).

Kvalitetsmått	Skillnad: denna grupp - normdata
Overt skönmålning	0.18
Kovert skönmålning	0.49**
Differentiering av svaren	0.46**
Jasägartendens	0.48**
Strukturlikhet	0.27

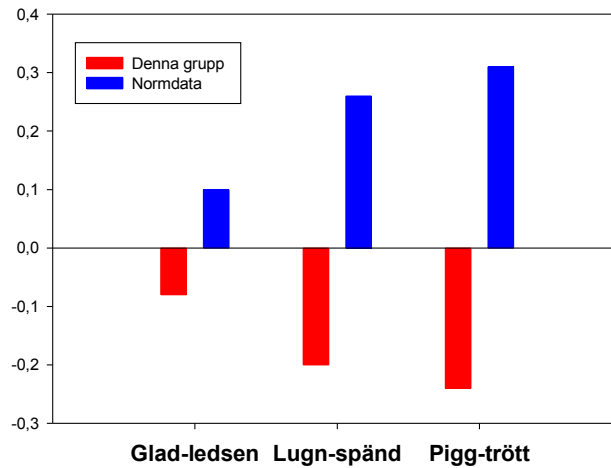
** p <0.01

Det fanns alltså en tendens att skönmåla svaren, och att bejaka positiva påståenden och förneka negativa påståenden. Höga värden på differentiering och strukturlikhet tyder samtidigt på att de testade lagt ner stor omsorg på att besvara testet. Dessa resultat tyder på att de testade uppfattade situationen ungefär som om de testats i skarpt läge, alltså som om deras resultat skulle vara av stor betydelse för dem. Eftersom arbetsgivaren inte skulle få del av

resultatet för enskilda testade kan detta ha varit en oberättigad tolkning. Å andra sidan skulle de själva få del av resultaten och företaget skulle få del av medelvärden för gruppen.

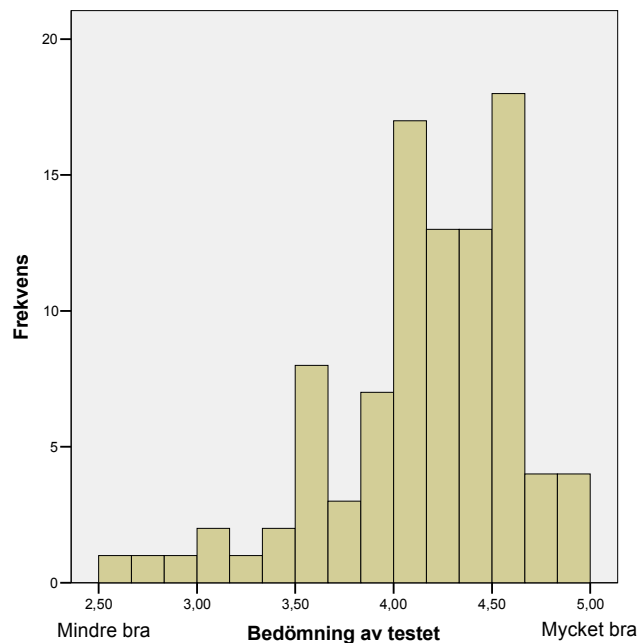
Stämningläget inför testningen var betydligt sämre i denna grupp än i normdata, se Fig. 11.2. Detta kan hänga samman med att testningen uppfattades som ”skarp”.

Stämningläge



Figur 11.2. Stämningläge inför testningen.

Bedömningen av testet framgår av Fig. 11.3.



Figur 11.3. Fördelning av bedömningarna av testets värde enligt dem som tog testet.

Figuren visar att de flesta gjorde en positiv bedömning av testet. Tabell 11.4 ger skillnaderna i medelvärden mellan denna grupp och normdata.

Tabell 11.4. Skillnader mellan denna grupp och normdata, standardiserade data.

Skala	Medelvärdeskillnad denna grupp - normdata	Skillnadens signifikans enligt <i>t</i> - testning
Utåtvändhet	0.16	
Vänlighet	0.03	
Emotionell stabilitet	-0.45	***
Öppenhet	0.40	***
Noggrannhet	0.12	
Uthållighet	0.05	
Samarbetsvilja	0.14	
Positiv grundhållning	-0.05	
Självförtroende	0.15	
Social förmåga	0.18	
Emotionell intelligens, självrapport	0.00	
Kreativitet	0.09	
Perfektionism	0.31	**
Arbetsstillfredsställelse	0.09	
Arbetsvilja	0.05	
Resultatorientering	0.50	***
Förändringsvilja	0.41	**
Arbetsintresse	-0.23	
Balans arbete – övrigt liv	0.04	

Resultaten visar:

- Skillnader mellan gruppen och normdata i personlighet endast i tre av skalorna: emotionell stabilitet (gruppen lägre), öppenhet och perfektionism (gruppen högre i dessa två skalor)
- Skillnader med avseende på resultatorientering och förändringsvilja till fördel för gruppen men också något lägre arbetsintresse.

Kriterievariablernas struktur

En komponentanalys gjordes av de 40 kriterievariablerna för att ge en översiktlig bild av data och samtidigt fånga upp de dominerande och viktigaste aspekterna. Fig. 11.4 visar ett s.k. scree-test [17] som används för att ta ställning till hur många komponenter (faktorer) som ger en god och rimlig approximation till data. Figuren visar att 2 komponenter ger en mycket god approximation, som förklarar 53.2% av den totala variansen i kriteriebedömningarna.

Laddningarna i de två komponenterna efter rotation (oblik) framgår av Tabell 11.5. Endast värden >0.4 eller <-0.4 har medtagits i tabellen. Komponenterna hade en svag positiv korrelation. Två kriterier har uteslutits eftersom de inte hade betydande laddning i någon av komponenterna. (Inte hinna bli klar, antalet avklarade uppgifter).

Resultatet är intressant på flera sätt:

- En mycket klar struktur framträder i data: en komponent mäter effektivitet i de centrala arbetsuppgifterna och en mäter social anpassning i jobbet. Detta stämmer bra med tidigare forskning.
- Tre kriterievariabler är laddade i båda komponenterna. Dessa mäter bedömarens syn på om den anställde är av sådant värde för företaget att man absolut vill behålla honom/henne.

För den fortsatta analysen skapades tre index som mäter effektivitet, social funktion och ”anställa och behålla”, kallad Värde för företaget (de tre komplexa kriterierna).

Reliabilitetsvärdena (Cronbachs alfa) för dessa kriterieindex var:

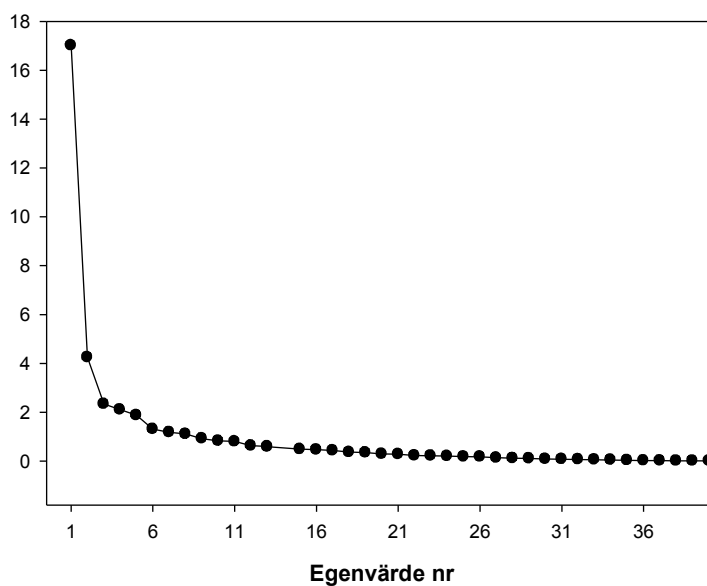
Effektivitet: 0.92, 10 skalor

Socialt fungerande: 0.93, 15 skalor

Värde för företaget: 0.93, 3 skalor

Effektivitet och socialt fungerande relaterades till bedömningen av värde för företaget. Den justerade multipla korrelationen (R^2_{adj}) var = 0.78. Betavikterna (standardiserade regressionsvikter) var 0.49 och 0.56 för effektivitet och socialt fungerande, respektive. De var ungefär lika.

Egenvärde



Figur 11.4. Scree-test för de 40 kriterievariablerna.

Tabell 11.5. Komponentladdningar, 2 komponenter och 38 kriterievariabler.

	Komponent 1	Komponent 2
Har god förståelse av arbetsuppgifterna	0.94	
Kvaliteten på hans/hennes arbetsinsats i stort	0.86	
Är kvalitetsmedveten	0.83	
Visar intresse för jobbet	0.80	
Arbetsinsatsen i sin helhet	0.77	
Löser problem som kan uppkomma i jobbet	0.76	
Är nogga med viktiga detaljer	0.75	
Anstränger sig för att göra ett bra jobb	0.74	
Anstränger sig	0.72	
Gör sig förtjänt av bonus	0.69	
Sätter ett värde på att uppnå resultat i jobbet	0.68	
Är engagerad i företagets resultat	0.67	
Arbetar energiskt	0.67	
Planerar sitt arbete på ett bra sätt	0.67	
Arbetar självständigt	0.66	
Visar en positiv utveckling av attityd, motivation och kompetens	0.60	0.40
Är den sortens medarbetare som vi gärna vill anställa och behålla i företaget	0.55	0.51
Är kreativ, kommer med bra idéer	0.54	
Har viljan att stanna på detta jobb	0.53	
Vi vill absolut behålla denna person i företaget	0.52	0.52
Är en effektiv medarbetare i vårt företag	0.52	
Arbetar gärna på övertid	0.47	
Ineffektiv tidsanvändning	-0.45	
Samarbetar gärna med andra		0.82
Omfattning av frånvaro från jobbet på grund av sjukdom		-0.80
Är hjälpsam i förhållande till kamrater		0.74
Är en trevlig och sympatisk person		0.72
Han/hon är positiv och omtyckt		0.69
Omfattning av frånvaro av okända skäl		-0.68
Är pålitlig		0.68
Det är lätt för chefer att ha att göra med honom/henne		0.68
Löser sociala problem på arbetsplatsen, t ex vid konflikter		0.68
Omfattning av frånvaro av andra legitima skäl		-0.65
Har en positiv attityd till företaget		0.59
Tar emot kritik på ett bra sätt		0.59
Låter också andra få del beröm eller belöningar för goda insatser		0.58
Klagar på arbetsvillkor och/eller lön		-0.56
Har en positiv attityd till kunder		0.44

Kriterier i förhållande till testresultat

Tabell 11.6 ger korrelationerna mellan testvariablerna och de tre sammansatta kriterievariablerna.

	Värde för företaget	Effektivitet	Social funktion
Utåtvändhet	0.39**	0.30*	0.38**
Vänlighet	0.24	0.06	0.27*
Emotionell stabilitet	0.40**	0.17	0.32*
Öppenhet	0.17	0.08	0.08
Noggrannhet	0.04	0.05	0.12
Uthållighet	0.45**	0.19	0.29*
Samarbetsvilja	0.37**	0.11	0.26
Positiv attityd	0.53**	0.20	0.35*
Självförtroende	0.31*	0.20	0.24
Social förmåga	0.14	-0.07	0.26
Emotionell intelligens	0.03	-0.23	0.11
Kreativitet	0.17	0.09	0.09
Perfektionism	-0.17	-0.10	-0.13
Arbetsstillfredsställelse	0.29*	0.12	0.15
Arbetsvilja	0.29*	0.07	0.18
Resultatorientering	0.32*	0.19	0.22
Förändringsvilja	0.23	0.02	0.20
Arbetsintresse	0.33*	0.33*	0.11
Balans	0.39**	0.20	0.31*
Samtliga testvariabler med lika vikter, perfektionism omvänt	0.34**	0.28*	0.20
Sex bästa testvariablerna, lika vikter	0.50**	0.39**	0.29*
Kovert skönmålning	0.27*	0.00	0.28*
Overt skönmålning	0.41**	0.17	0.31*
Variationsvidd i korrelationer	-0.17 till +0.53	-0.10 till +0.33	-0.13 till +0.38
Median	0.29	0.11	0.24

*p < 0.05

** p<0.01

Det är slående att korrelationerna varierar starkt. De låga värdena, t ex med kreativitet, framstår som naturliga med tanke på arbetsinnehållet. Genomsnittet eller medelvärdet av korrelationerna är inte ett särskilt meningsfullt sätt att beskriva testets värde. En multipel korrelation går inte att beräkna med N=53 och antalet oberoende variabler=19, men

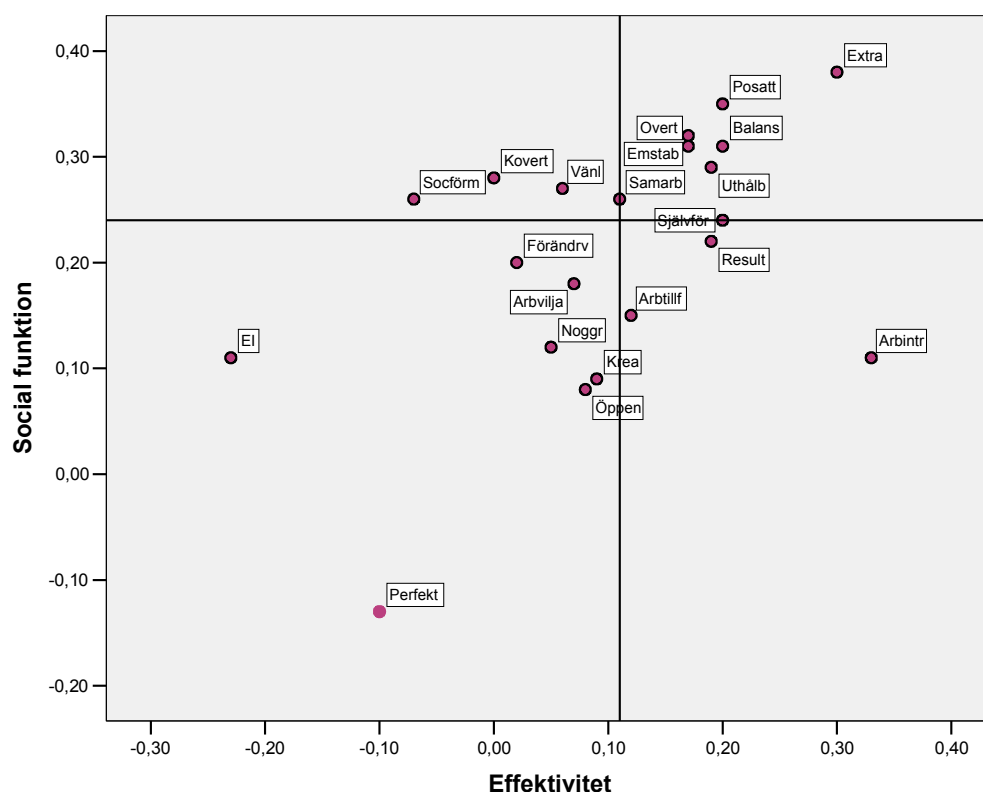
korrelationerna med indexen där alla testvariabler ingår med lika vikt är informativa: 0.33, 0.28 och 0.20.

Det bedömda värdet för företaget tycks främst hänga samman med nedanstående testvariabler:

- Positiv attityd
- Uthållighet
- Emotionell stabilitet
- Utåtvändhet
- Samarbetsvilja
- Balans

Dessa testvariabler är av generellt värde även för de andra kriterierna. Ett sammanslaget index med dessa de sex bästa testvariablerna med lika vikter gav som framgår av Tabell 11.6 korrelationerna 0.50, 0.39 och 0.29 för kriterierna Värde för företaget, Effektivitet och Social funktion. Dessa värden ger en betydligt mera informativ bild av testets validitet än mediankorrelationerna.

Hur de olika testvariablerna fungerar gentemot kriterierna åskådliggörs också i Fig. 11.5, där kolumnerna 3 och 4 i Tabell 11.6 har prickats mot varandra. De heldragna strecken ger positionerna för medianvärdena. Figuren visar att de två kriterierna fungerar relativt likartat gentemot testdimensionerna, även om effektivitet har lägre korrelationer än social fungerande. Vänlighet och social förmåga ligger relativt högt mot social funktion och lågt mot effektivitet, vilket är rimligt utifrån en matchningshypotes, som säger att variabler med likartat innehåll har särskilt stor chans att korrelera högt [94]. Det motsatta gäller resultatorientering, arbetsintresse och arbetstillfredsställelse. Åtskilliga testdimensioner tycks emellertid fungera likartat för de två typerna av kriterier.



Figur 11.5. Prickning av kriteriekorrelationerna, social funktion mot effektivitet.

Proxykriterier

Tabell 11.7 visar korrelationerna mellan personlighetsskalorna och de 6 proxykriterierna.

Tabell 11.7. Korrelationer mellan personlighetsvariabler och proxykriterier. De två största värdena i varje kolumn, bortsett från skönmålning, i fetstil.

	Arbets- tillfreds- ställelse	Arbets- vilja	Resultat- orienteri- ng	Föränd- ringsvilj- a	Arbets- intresse	Balans	Genom- snittlig korrela- tion
Utåtvändhet	0.03	0.15	0.48**	0.27*	0.23	0.12	0.21
Vänlighet	0.33*	0.33*	0.30*	0.39**	0.28*	0.44**	0.35
Emotionell stabilitet	0.20	0.31*	0.21	0.20	0.16	0.50**	0.26
Öppenhet	0.28*	0.30*	0.42**	0.59**	0.41**	0.31*	0.39
Noggrannhet	0.17	0.22	0.26	0.09	0.17	0.42**	0.22
Uthållighet	0.23	0.26	0.33*	0.51**	0.36**	0.44**	0.36
Samarbetsvil- ja	0.31*	0.37**	0.48**	0.57**	0.28*	0.55**	0.43
Positiv attityd	0.75**	0.80**	0.67**	0.67**	0.63**	0.58**	0.68
Självförtroen- de	0.28*	0.40**	0.70**	0.49**	0.48**	0.51**	0.48

Tabell 11.7. Korrelationer mellan personlighetsvariabler och proxykriterier. De två största värdena i varje kolumn, bortsett från skönmålning, i fetstil.

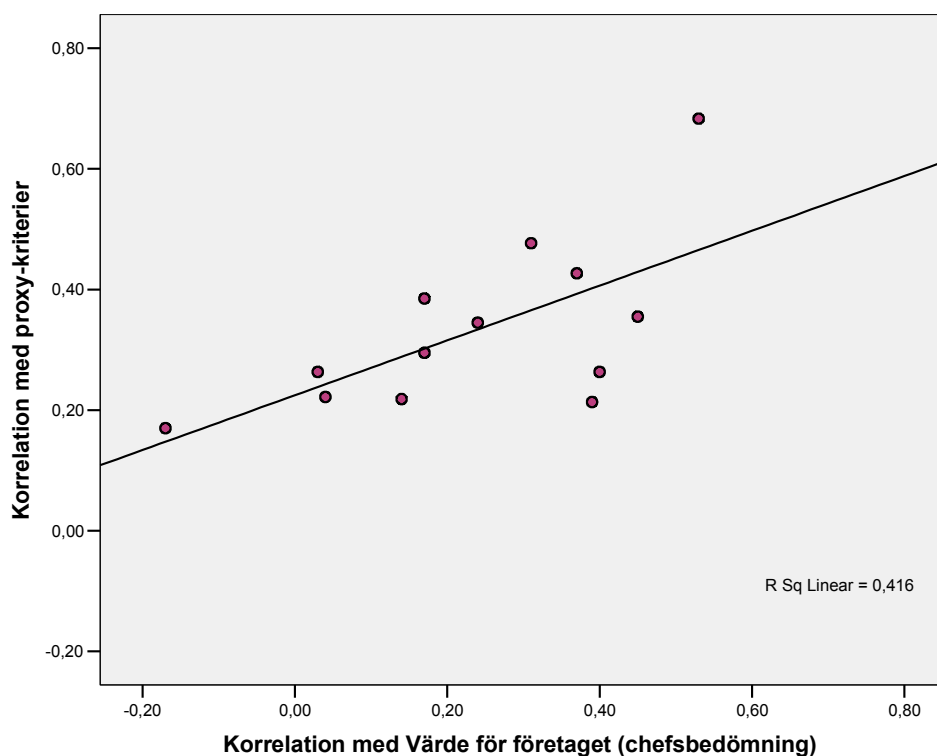
Social förmåga	0.03	0.12	0.38**	0.41**	0.09	0.28*	0.22
Emotionell intelligens	0.13	0.26	0.36**	0.33*	0.11	0.39**	0.26
Kreativitet	0.31*	0.26	0.22	0.43**	0.51**	0.04	0.30
Perfektionism	0.20	0.22	0.19	0.11	0.19	0.11	0.17
Median	0.23	0.26	0.36	0.41	0.28	0.42	
Kovert skönmålning	0.47**	0.25	0.45**	0.47**	0.38**	0.57**	
Overt skönmålning	0.35**	0.59**	0.35*	0.29*	0.29*	0.39**	

*p<0.05

** p<0.01

Tabell 11.7 visar höga samband mellan personlighetsvariablerna och proxykriterierna. Positiv attityd framstår som den viktigaste dimensionen, tätt följd av självförtroende. I endast ett par fall visade ”big five”-skalor bättre validitet än *UPP*-testets unika skalor.

Den genomsnittliga korrelationen mellan personlighetsskalorna och proxykriterierna beräknades. Denna relaterades till personlighetsskalornas validiteter mot chefsbedömningarna. Det visade sig då att den genomsnittliga proxyvaliditeten korrelerade mycket högt med de samband som erhållits med Värde för företaget (0.64), lägre men fortfarande ganska högt med Effektivitet (0.39) och Social funktion (0.43). Detta resultat är mycket viktigt eftersom det tyder på att proxyvalidering kan förväntas ge resultat som liknar dem som erhålls genom validering mot förmansbedömningar, i synnerhet för ett övergripande kriterium som Värde för företaget. Fig. 11.6 ger en prickning av de två serierna av korrelationer mot varandra.



Figur 11.6. Prickning av validitet enligt proxykriterierna mot validitet i förhållande till förmansbedömningar.

Om detta resultat står sig i nya undersökningar - vilket är rimligt att förvänta - öppnas möjligheten till enklare och billigare testvalideringar, vilket i sin tur borde kunna uppmuntra testkonstruktörer och -leverantörer av utländska test - att pröva värdet av sina produkter. *UPP*-testet finns också i form av en modul som enbart mäter proxykriterier, den kan kopplas till vilket test eller vilken annan bedömningsprocedur som helst.

Sammanfattning och slutsatser

UPP-testet visade sig ha goda psykometriska egenskaper i denna studie. Det bedömdes också positivt av de allra flesta av dem som tog testet. Formuläret för bedömning av arbetsresultat visade sig ha en tydlig 2-dimensionell struktur i enlighet med förväntningarna.

Resultaten visade att

- Den testade gruppens svar tydde på att de upplevde situationen som ”skarpt läge”: de hade en tendens att skönmåla men också en tydlig tendens att ge genomtänkta svar
- Gruppen utmärktes av bara några få personlighetsskillnader gentemot normgruppen. Den lägre emotionella stabiliteten kan bero på demografiska faktorer eller stämningläget
- Den testade gruppen visade höga värden i resultatorientering och förändringsvilja, men något lägre än normgruppen i arbetsintresse
- Validering mot kriterierna Värde för företaget, Effektivitet och Social funktion fungerade bra, särskilt för kriteriet Värde för företaget

- Detta kriterium var relaterat till positiv attityd, utåtvändhet, emotionell stabilitet, samarbetsvilja och uthållighet
- Sambanden var starkare med proxykriterierna, men dessa samband var mycket likartade dem som erhålls vid validering mot de av testningen helt oberoende förmansbedömningarna

Social anpassning i arbetsmiljön

Denna studie byggde på samma material som i den föregående undersökningen av chefsansvar. Data kompletterades med:

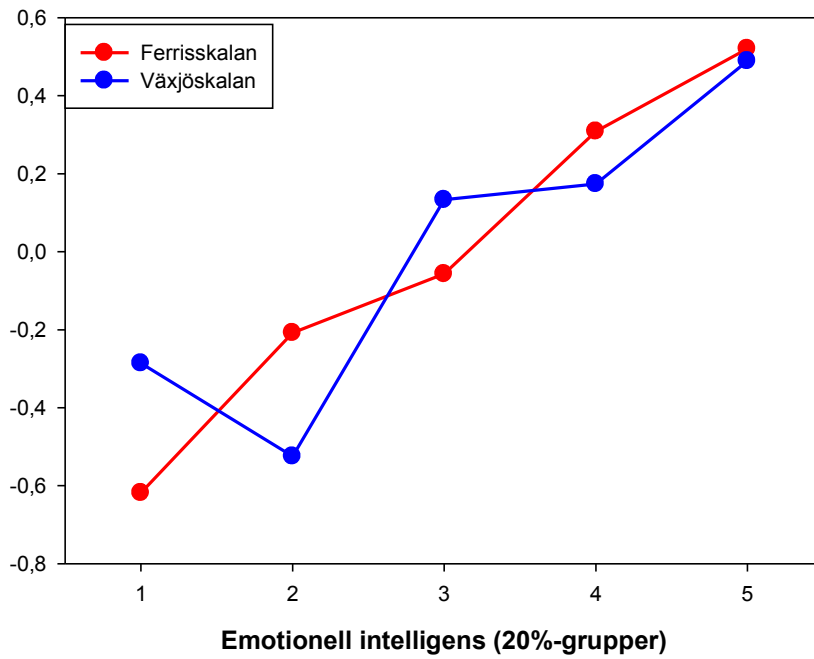
- En skala för mätning av social anpassning i arbetet. Denna skala är konstruerad av fil. dr. Kristiina Möller vid Linnéuniversitetet och kommer att här kallas Växjö-skalan.
- Ferris skala [30] för mätning av social arbetsanpassning.

Reliabilitetsvärden för de analyserade skalorna återges i Tabell 11.8.

Variabel	Reliabilitet (alfa-koefficient)
Ferrisskalan	0.73
Växjöskalan	0.95
UPP: EI självrapport	0.73
UPP: EI bildskalan	0.91
UPP: social förmåga	0.71

Fig. 11.8 visar sambandet mellan social anpassning och emotionell intelligens. I termer av korrelation på individnivå var sambandet = 0.37, $p < 0.0005$ för det sammanslagna måttet.

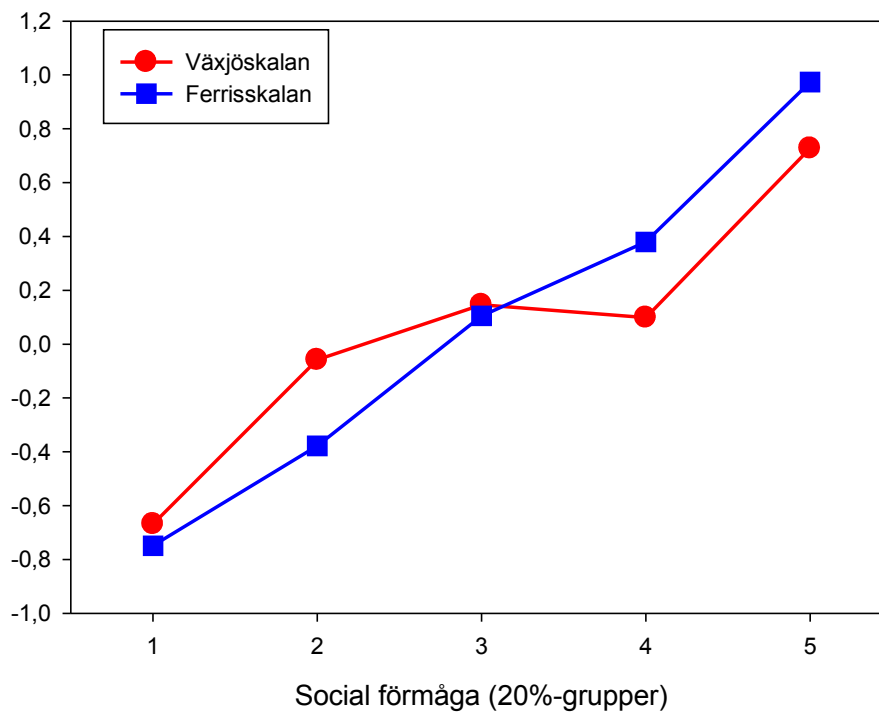
Social anpassning i arbetsmiljön



Figur 11.8. Sambandet mellan emotionell intelligens och social anpassning i arbetet.

Fig. 11.9 ger motsvarande resultat för UPP-testets mått på social förmåga. Korrelationen var = 0.58, $p < 0.0005$ för det sammanslagna måttet.

Social anpassning i arbetsmiljön



Figur 11.9. Sambandet mellan social förmåga och social anpassning i arbetet.

Det sammansatta måttet på social arbetsfunktion har slutligen analyserats i förhållande till testdata med hjälp av hierarkisk multipel regression. I denna analys prövades först femfaktormodellen, sedan EI och social förmåga enligt *UPP*-testets tillkommande variabler. Se Tabell 11.9.

Tabell 11.9. Hierarkisk regressions analys av social arbetsfunktion					
Grupp av oberoende variabler	R^2_{adj}	ΔR^2	F för ΔR^2	df	p
Steg 1: FFM-modellen	0.105	-	-		0.006 (FFM-modellen)
Steg 2: FFM + EI och social förmåga	0.323	0.221	17.320	2,99	<0.0005 (tillskottet i steg 2)

Kriteriekorrelationen som svarar mot 32.3 % förklarad varians är = 0.57. Tillskottet som *UPP*-testets relevanta variabler svarar för utöver FFM är mycket stort och starkt signifikant.

Beta-värdena i regressionsekvationen var inte signifikanta för någon av variablerna i FFM, men däremot tydliga för social förmåga och i viss mån för EI, se Tabell 11.10.

Tabell 11.10. Social anpassning i arbetsmiljön som en funktion av personlighet. Beta-koefficienter för regressionsmodellen i steg 2.		
Variabel	Beta-värde	p
Vänlighet	0.071	0.417
Öppenhet	-0.062	0.529
Emotionell stabilitet	-0.012	0.888
Extraversion	-0.003	0.974
Noggrannhet	-0.029	0.743
Emotionell intelligens	0.171	0.077
Social förmåga	0.520	0.000

UPP-testets mått på EI och social förmåga hade alltså hög validitet i denna studie.

Arbetsprestation (självbedömd)

Arbetsprestation (självbedömd) samt arbetsmotivation och andra arbetsrelaterade attityder har använts som kriterier. Arbetsprestationen studerades på ett tidigt stadium av utvecklingsarbetet (Studie 1) och gav samband med testvariablerna omkring 0,50 (multipel

korrelation, alla värden korrigerade för skönmålning). Se Tabell 11.11. En åtskillnad gjordes mellan kärnuppgifter och sociala dimensioner [74]: validiteten var likvärdig i de två fallen.

Tabell 11.11. Validitetskoefficienter. De två högsta värdena i varje kolumn i fetstil.

	Arbets- prestation, kärn- uppgifter, korrigerad	Arbets- prestation, sociala dimen- sioner, korrigerad	Arbets- prestation, global, korrigerad	Arbets- prestation, global, okorri- gerad
Extraversion korrigerad	0,40	0,46	0,24	0,41
Social förmåga korrigerad	0,31	0,43	0,18	0,35
Självförtroende korrigerad	0,45	0,43	0,36	0,43
Emotionell stabilitet korrigerad	0,07	0,13	0,18	0,10
Noggrannhet korrigerad	0,16	0,08	0,17	0,12
Uthållighet korrigerad	0,29	0,22	0,32	0,25
Positiv grundhållning korrigerad	0,28	0,31	0,13	0,29
Emotionell intelligens korrigerad	0,27	0,16	0,19	0,21
Passiv aggression korrigerad	-0,27	-0,18	-0,17	-0,24
Jagstyrka korrigerad	0,52	0,46	0,39	0,39
Jagstyrka ej korrigerad	0,36	0,32	0,27	0,31
Multipel korrelation med okorrigerade kriterier (justerad):	0,49	0,50	0,36	0,72

Jagstyrka, korrigerad för skönmålning, hade högsta validiteten, därefter kom självförtroende och sedan extraversion.

Omfattande forskning tyder på att självbedömningar är användbara som mått på prestation [2; 41]. Ett visst värde tycks alltså vara klart dokumenterat för ”subjektiv” framgång; detta gäller inte bara individer utan även på organisationsnivå [121]. Henemans tidiga studie [41] är intressant eftersom han visade att självskattningar fungerade bra i en situation där inget av stor vikt stod på spel för de testade – detta är ju fallet i den studie som rapporteras här.

Självbedömd arbetsprestation är emellertid en problematisk variabel eftersom det är väl känt att de flesta systematiskt överskattar sin egen prestation [16; 26]. Detta är emellertid en fråga om nivå på bedömningen och om alla har ungefär samma tendens att lägga till ett visst positivt tillskott så behöver inte variabeln som sådan vara mindre informativ, när det kommer till variationer mellan individer. Det är också intressant att notera att problemet tycks vara störst när det gäller att bedöma, och minnas, sina mindre lyckade prestationer.

Data på självbedömning visar alltså att det finns en viss information i den och våra egna resultat tyder på samma sak.

Arbetsmotivation och övriga arbetsrelaterade variabler (proxy-kriterier)

En enkel metod för validering är att använda proxykriterier, alltså data som kan fungera som ersättning för oberoende bedömningar och som kan samlas in i samband med testningen. Det kan handla om mått på arbetsmotivation eller balans mellan arbete och övrigt liv. Det är väl känt från mycket omfattande forskning att sådana data är relaterade till arbetsresultat. Se t ex Judge m fl [54] för en metaanalys av arbetstillfredsställelse och arbetsresultat. Stöd för proxyvalidering har getts ovan, se Fig. 11.6.

Resultat av proxyvalidering

Omfattande valideringsarbete genomfördes mot arbetsmotivation och övriga arbetsrelaterade variabler. Här redovisas endast resultat från normmaterialet men analyser av övriga studier gav liknande resultat. Multipla korrelationer för vart och ett av de sex kriterierna mot FFM-modellen Big Five och mot *UPP*-testet framgår av Tabell 11.12. I *UPP*-testet ingick också de 8 tillkommande variablerna – uthållighet, samarbetsvilja etc.

Tabell 11.12 Validering mot 6 kriterievariabler, förklarad varians (kvadrerade multipla korrelationer).

Kriterievariabel	Big Five	<i>UPP</i> -testet
Förändringsvilja	0,137	0,274
Arbetstillfredsställelse	0,008	0,454
Arbetsvilja	0,010	0,475
Resultatorientering	0,106	0,212
Arbetsintresse	0,054	0,389
Balans arbete – övrigt liv	0,045	0,097
Medelvärde	0,041	0,317

I samtliga fall var ökningen av validiteten högt signifikant och av dramatisk storlek. Förklarad varians på 0,317 motsvarar en korrelation på 0,56. Märk att det var balansdimensionen som drog ner värdet för *UPP*-testet i viss mån. Övriga kriterier kunde förklaras på högre nivå.

För Big Five var den genomsnittliga korrelationen med kriterierna endast 0,20, ett ganska normalt värde för personlighetstest. En ökning av validiteten från 0,20 till 0,56 kan påvisas ge mycket stora personalekonomiska vinster om testet används vid urval.

Dessa värden på validiteten påverkas ej av skönmålning eller andra felfaktorer eftersom dessa rensats bort ur både beroende och oberoende variabler. Om man inte gör denna korrektion blir validiteten, i termer av korrelation mellan test och kriterium, förhöjd med ca 0,10, vilket ger en överskattning av testets värde.

Tabell 11.13 ger relativa vikter beräknade enligt Johnsons metod [48-50]. De traditionella beta-vikterna i multipel regressionsanalys kan vara vilseledande eftersom de är starkt

beroende av kontexten av olika förklarande eller oberoende variabler. En variabel av betydelse kan få en alltför låg beta-vikt om den korrelerar med en annan oberoende variabel. Johnsons metod uppskattar varje oberoende variabels bidrag till förklaringen av den beroende variabeln, både det unika bidraget och bidraget tillsammans med andra oberoende variabler.

Tabell 11.13. Relativa vikter för testvariablerna vid förklaring av kriterierna. Varje vikt ger den andel av det totala förklaringsvärdet som testvariabeln svarar för. De två största värdena i varje kolumn anges i fetstil.

	Föränd- ringsvilja	Arbets- tillfreds- ställelse	Arbets- vilja	Balans	Arbets- intresse	Resultat- orientering
Vänlighet	1,3	1,7	1,9	5,9	1,3	1,6
EI	0,9	0,3	3,9	4,6	3,3	2,4
Öppenhet	12,3	0,5	3,8	1,7	2,4	2,3
Emotionell stabilitet	2,8	0,3	2,2	2,3	2,8	9,6
Extraversion	3,7	25,9	6,0	1,0	1,5	3,7
Perfektionism	20,0	0,9	8,9	2,9	3,8	12,0
Kreativitet	12,5	0,3	9,8	64,9	10,4	25,3
Uthållighet	1,5	0,2	0,5	5,1	66,6	1,7
Noggrannhet	2,7	2,2	9,0	1,4	0,9	1,0
Samarbetsvilja	3,5	65,6	28,3	7,1	0,5	18,9
Positiv attityd	24,9	1,4	2,9	1,3	5,4	10,6
Självförtroende	13,2	0,5	,6	1,3	0,9	10,7
Social förmåga	0,7	0,2	0,1	0,5	0,3	0,2

En övergripande bedömning visar att de viktigaste personlighetsdimensionerna var:

1. Samarbetsvilja
2. Kreativitet
3. Extraversion
4. Uthållighet
5. Positiv attityd
6. Perfektionism

Resultaten avseende validering torde vara mycket stabila eftersom stickprovet var stort. Det är intressant att av de sex viktigaste variablerna var det endast en, extraversion, som ingår i FFM. I båda de fall där extraversion var bland de två bästa förklarande variablerna (arbetstillfredsställelse och arbetsvilja) en variabel bland dem som tillkom utöver FFM ännu bättre (samarbetsvilja).

Livsframgång

Inkomst

Samma grupp som användes i analysen av chefsansvar tillfrågades även om sin inkomst, i kronor och per månad. Antalet svarande var 107. *UPP*-testets variabler grupperades (Big Five, *UPP*-testets tillkommande variabler samt proxykriterier) och relaterades till inkomst i en linjär multipel regression. Den justerade multipla korrelationen var 0,100 vilket svarar mot en korrelation på 0,31. $p=0.008$. Med andra ord uppnåddes en tydlig validitet mot detta viktiga (externa) kriterium på livsframgång, med hög statistisk signifikans.

Uppnådd utbildningsnivå

Uppnådd utbildningsnivå är en konsekvens bl a av intellektuell förmåga, uthållighet, social förmåga, etc. *UPP*-testet är orelaterat till intellektuell förmåga, och ett samband med uppnådd utbildningsnivå borde därför på ett ”rent” sätt avspegla testets externa validitet för en mycket viktig aspekt på livsframgång. Ålder är emellertid ännu en faktor som kan påverka resultaten – ju äldre man är desto större har chansen varit att fullfölja utbildning - och den kommer att analyseras i det följande.

Den totala datamängden av testningar med hela *UPP*-testet innehöll vid analystillfället resultat från 1111 personer. Uppnådd utbildningsnivå användes som beroende variabel i en multipel regressionsanalys i vilken de oberoende variablerna var FFM-skalorna, *UPP*-testets tillkommande skalor samt proxykriterierna. Hela datamängden analyserades, liksom män och kvinnor separat. Se Tabell 11.14 för resultaten, som ger korrelationerna (kvadratroten ur R^2_{adj}).

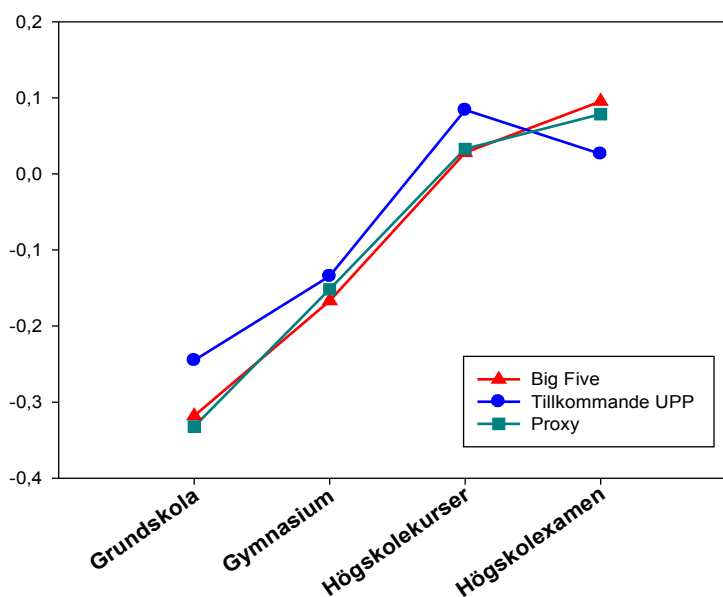
Tabell 11.14 Korrelationer mellan *UPP*-testet och uppnådd utbildningsnivå.

Grupp	
Hela materialet	0,26***
Män	0,30***
Kvinnor	0,22***

*** $p < 0,0005$

Trenderna var mycket likartade för FFM-skalorna, tillkommande variabler i *UPP*-testet och proxykriterierna, se Fig. 11.10.

Medelvärde, standardiserad skala



Figur 11.10. Medelvärden i grupper av testvariabler som funktioner av uppnådd utbildningsnivå.

I termer av d -värden är sambanden som framgår av figuren medelstarka, omkring 0,3 – 0,4. Ålder kunde, som nämndes ovan, vara en förklaring till dessa samband eftersom äldre personer har haft mera tid för att genomföra utbildning. I materialet var medianåldern 27 år, och variationen mycket stor (17-76). Eftersom variationen var så stor i ålder borde denna faktor ha en god chans att påverka resultaten. För att undersöka denna möjliga felkälla gjordes en separat regressionsanalys av utbildningsnivå mot ålder och residualen togs som mått på utbildning "rensat" från ålderseffekten. Korrelationen i hela materialet mellan utbildningsnivå och ålder sjönk då, men bara obetydligt, till 0.26.

Sammanfattning och slutsatser om validering

I detta avsnitt sammanfattas resultaten av valideringarna, både begreppsvalideringar och valideringar mot externa kriterier. Eftersom *UPP*-testet skall utvärderas mot europeisk standard för test enligt EFPA²⁰ citeras det grundläggande EFPA-dokumentet²¹ [59].

²⁰ European Federation of Psychologists' Associations

²¹ <http://www.efpa.eu/professional-development/tests-and-testing>

Begreppsvalideringar, enligt EFPA:

”Construct validity includes correlations of scales from similar instruments.” Sid. 17 [59].

Kriterierelaterade valideringar, enligt EFPA:

“Concurrent and predictive validity refer to studies where real-world criterion measures (i.e. not other instrument scores) have been correlated with scales”. Sid. 17 [59].

Sammanvägning av båda typerna av validitet, enligt EFPA:

“When judging overall validity, it is important to bear in mind the importance placed on construct validity as the best indicator of whether a test measures what it claims to measure” sid. 18 [59].

Kommentar

I detta avsnitt har redogjorts för omfattande valideringar av båda typerna. Olika typer av kriterier har använts. Resultaten är goda. Kanske kan viss tveksamhet föreligga när det gäller stickprovets storlek i vissa fall eftersom man kan tolka EFPA:s kriterier som ett krav på att stickprovet måste omfatta minst 100 personer. Frågor om detta till ordföranden i EFPA-gruppen har visat att det inte rör sig om ett krav utan en rekommendation. Om mindre stickprov ger signifikanta resultat är dessa relevanta²².

En innovation i vårt arbete är proxyvalidering. Den har gett mycket uppmuntrande resultat. Det är också viktigt att data tyder på att den kan ge resultat fullt jämförbara med validering mot ett externt kriterium som förmansbedömningar. Här öppnas en möjlighet till snabba valideringar till låg kostnad; den borde vara av allmänt intresse eftersom alla test kan kopplas till den modul som mäter proxykriterierna.

EFPA:s riktlinjer säger att begreppsvalideringen är mycket viktig, viktigare än kriterievalideringen. Som framgår av detta avsnitt har samtliga testvariabler begreppsvaliderats, med goda resultat med bara ett undantag (perfektionism). I det fallet är kriteriet av tveksamt värde och fortsatt forskning påkallad.

UPP-testet innehåller komponenter som vanligen inte finns i personlighetstest, nämligen mätningar av

1. Stämningssläge
2. Datakvalitet
3. Attityder till testet

Reliabiliteten hos dessa skalor har redovisats, liksom begreppsvaliditeten hos datakvaliteten. Stämningsslägesmätningen bygger på väl etablerade och ofta studerade metoder [96; 116].

12. Intellektuell förmåga

Sambandet mellan testvariablerna och högskoleprovet var omkring 0, se Tabell 12.1, som bygger på Studie 1 i den tekniska rapporten. Detta är ett önskat resultat eftersom UPP-testet avses ge kompletterande information utöver intellektuell förmåga.

²² Se denna länk: <http://www.psykologisk-metod.se/files/Bartram%202.pdf>

Tabell 12.1. Korrelationer mellan testvariabler och högskoleprovet.

Testvariabel	Högskoleprovet
Extraversion korrigerad	-0,15
Social förmåga korrigerad	-0,16
Självförtroende korrigerad	-0,04
Emotionell stabilitet korrigerad	0,26
Noggrannhet korrigerad	-0,09
Uthållighet korrigerad	0,05
Positiv grundhållning korrigerad	-0,22
Emotionell intelligens korrigerad	0,06
Passiv aggression korrigerad	0,13
Jagstyrka korrigerad	-0,07
Jagstyrka ej korrigerad	-0,05
Multipel korrelation med okorrigerade kriterier (justerad):	0.00

Högskoleprovet är användbart som mått på intellektuell kapacitet men en direkt mätning med ett färdighetstest är bättre. I en särskild studie (studie 12) togs testet av 128 studenter (ekonomutbildning) i Lund, 59 män och 60 kvinnor. (Uppgift om kön saknas för 9 deltagare). Åldern varierade mellan 19 och 30 år, median 22 år. De tog även ett analogitest med 18 uppgifter, 5 svarsalternativ för varje uppgift, reliabilitet (alfa) = 0,68²³. Korrelationerna mellan testresultatet i analogitestet (antal rätt lösta uppgifter) och personlighetsvariablerna ges i Tabell 12.2, motsvarande resultat för arbetsrelaterade variabler i Tabell 12.3

²³ Gruppen bestod av intellektuellt högpresterande personer, varför en hög reliabilitet inte kan förväntas.

Tabell 12.2. Korrelationer mellan verbal/analytisk förmåga och personlighetsvariabler.

Testvariabel (korrigerade för skönmålning)	Korrelation
Vänlighet	-0,09
Emotionell intelligens	0,04
Öppenhet	0,27**
Emotionell stabilitet	-0,02
Extraversion	0,12
Perfektionism	-0,11
Kreativitet	0,18*
Uthållighet	-0,14
Noggrannhet	-0,08
Positiv hållning	-0,06
Självförtroende	-0,07
Social förmåga	-0,08
Jagstyrka	0,02
Stresskänslighet	-0,03

* $p < 0,05$

** $p < 0,01$

Endast öppenhet och kreativitet visade sig ha något samband med personlighet, men även dessa samband var svaga.

Tabell 12.3. Korrelationer mellan arbetsrelaterade variabler och verbal/analytisk förmåga.

Testvariabel (korrigerad)	Korrelation
Arbetsvilja	0,09
Arbetsintresse	0,21*
Arbetsstillfredsställelse	0,12
Resultatorientering	0,20*
Förändringsvilja	0,25**
Balans	-0,07

* $p < 0,05$

** $p < 0,01$

Även dessa samband var svaga, men kanske något större än sambanden med personlighetsvariablerna. Det låga sambandet mellan intellektuell förmåga och testet har alltså i och med denna studie dokumenterats även mot ett färdighetstest. Se även avsnitt 19.

13. Narrativt utlåtande

Dispositionen av det narrativa utlåtandet återges i Tabell 13.1.

Tabell 13.1. Rubriker i det narrativa utlåtandet och motsvarande testvariabler, i den ordning de rapporteras i utlåtandet

Rubrik i utlåtandet	Testvariabler	Texten bygger på
Sociala funktioner	Social förmåga, emotionell intelligens, samarbetsvilja, utåtvändhet, vänlighet	Stanine-värde
Emotionell anpassning	Emotionell stabilitet, uthållighet, positiv grundattityd, självförtroende	Stanine-värde
Noggrannhet	Noggrannhet, perfektionism	Stanine-värde
Kreativitet	Kreativitet, öppenhet	Stanine-värde
Jagstyrka	Sammansatt variabel: jagstyrka	Stanine-värde
Stresskänslighet	Sammansatt variabel: Stresskänslighet	Stanine-värde
Chefpotential	Sammansatt variabel: Chefpotential	Stanine-värde
Arbetsmotivation	Arbetsstillfredsställelse, arbetsvilja, arbetsintresse	Stanine-värde
Resultatorientering	Resultatorientering	Stanine-värde
Förändringsvilja	Förändringsvilja	Stanine-värde
Balans i livet	Balans	Stanine-värde
Övergripande bedömning av datakvalitet	Sammansatt mått på datakvalitet med lika vikter för nedanstående variabler	Stanine-värde
Skönmålning	Medelvärde av overt och kovert skönmålning	Stanine-värde
Ja-sägartendens	Tendensen att hålla med om påståenden i testet oavsett deras innehåll (medelvärdet av svaren på alla tesuppgifter utom bild-delen)	Stanine-värde
Differentiering mellan testuppgifter:	Tendensen att ge olika svar på olika uppgifter i testet i stället för att välja samma svarsalternativ för många av uppgifterna (den intra-individuella standardavvikelsen över samtliga uppgifter utom bild-delen)	Stanine-värde
Strukturlikhet	Den intra-individuella korrelationen mellan svaren på alla uppgifter utom bild-delen och medelvärdet för dessa uppgifter i normdata	Stanine-värde
Obesvarade uppgifter	Antalet ej besvarade uppgifter	Råvärde
Stämningsläge vid testningen	Medelvärdet av 12 frågor om stämningsläge, högre värde betyder mera positivt stämningsläge	Stanine-värde
Attityd till testet: "Face Validity":	Medelvärdet av 7 frågor som mäter attityden till testet, högre värde betyder mera positiv bedömning av testet	Stanine-värde

14. Jämställdhet och mångfald: rättviseaspekter

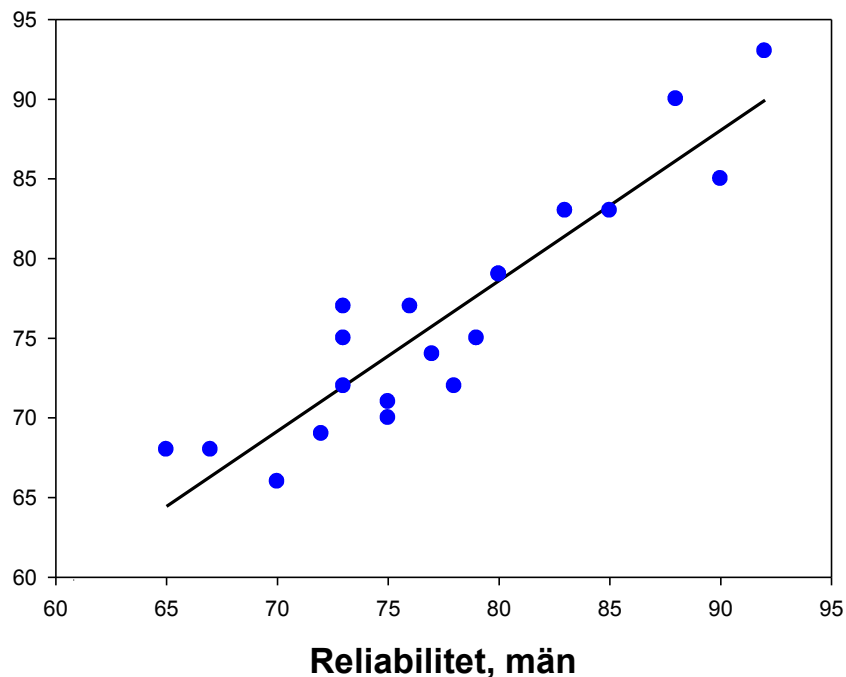
Jämställdhet

Kravet på alla test är att de inte ska gynna män framför kvinnor, eller tvärtom. Här kommer den frågeställningen att undersökas med avseende på testets reliabilitet och på skönmålning.

Resultat

Könsskillnader förekommer i vissa av testets variabler, t ex emotionell intelligens. Om man antar att denna skillnad är reell, dvs. inte uttryck för någon artefakt som beror på att män och kvinnor uppfattar testuppgifterna och svarsskalorna olika, finns ingen anledning att normera bort den genom att arbeta med separata normer för män och kvinnor. Detsamma torde gälla samtliga skalor, med tanke på hur likartat de fungerar när det gäller homogenitet (reliabilitet). Se Fig. 14.1.

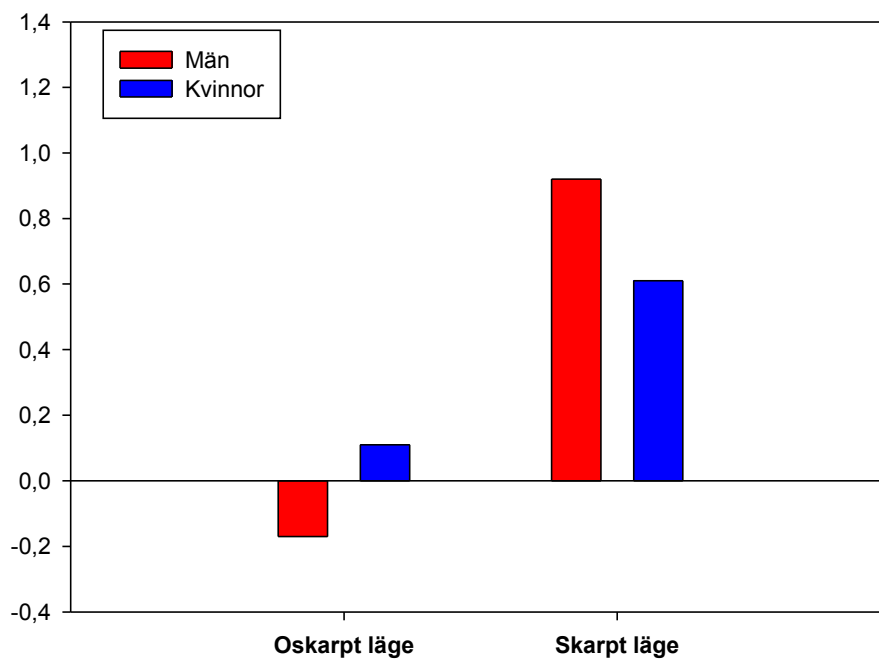
Reliabilitet, kvinnor



Figur 14.1. Reliabilitet (alfa-värden) för *UPP*-testets grundskalor, separat för män och kvinnor, normdatamaterialet.

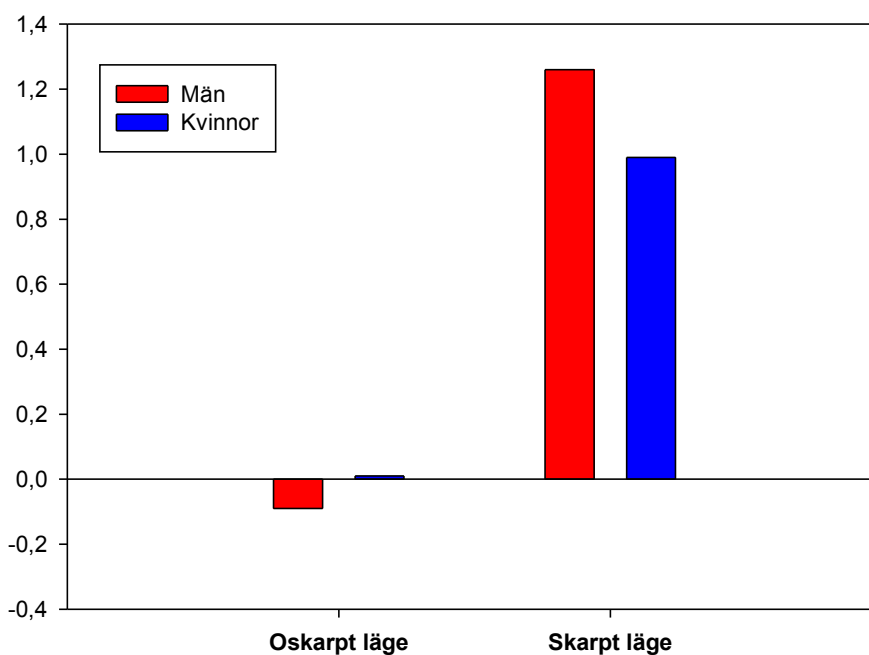
Män och kvinnor hade ungefär samma tendens att skönmåla i oskarpt läge, men i skarpt läge, vid ansökan om chefsbefattning, uppkom en tydlig skillnad. Männerna hade en starkare tendens att skönmåla, se Fig. 14.2 och 14.3. Figurerna visar också den väntade, och stora, skillnaden mellan oskarpt och skarpt läge i skönmålning.

Overt skönmålning



Figur 14.2. Könsskillnad i overt skönmålning, oskarpt och skarpt läge.

Kovert skönmålning



Figur 14.3. Könsskillnad i kovert skönmålning, oskarpt och skarpt läge.

Att männen skönmålar mera än kvinnor i skarpt läge gör att de vinner en fördel i testningen. Se Tabell 14.1 som visar hur bedömningen av chefslämplighet föll ut för män och kvinnor i skarpt och oskarpt läge, före och efter korrektion för skönmålning.

Tabell 14.1. Medelvärden i chefpotential före och efter korrektion för skönmålning, standardiserade värden.

	Före korrektion		Efter korrektion	
	Män	Kvinnor	Män	Kvinnor
Oskarpt läge	-0.05	-0.02	0.04	-0.05
Skarpt läge	1.00	0.86	-0.05	0.04

Tabellen visar:

- Före korrektion förelåg en stor skillnad mellan skarp och oskarpt läge
- Denna skillnad eliminerades tack vare korrektionen för skönmålning
- En könsskillnad till männens fördel i skarpt läge, före korrektion, eliminerades tack vare korrektionen.

Slutsatser om jämställdhet

Sammanfattningsvis kan den slutsatsen dras att UPP-testet fungerar psykometriskt väl och på samma nivå för män och kvinnor, samt att korrektionen för skönmålning skapade rättvisa mellan könen, en rättvisa som inte fanns i de okorrigerade data.

Mångfald

Det är väl känt att invandare har svårt att få jobb. Det kan finnas många orsaker till det. Användningen av psykologiska test borde kunna vara till fördel för invandrare som behärskar svenska eftersom ovidkommande faktorer som utseende och namn inte påverkar testresultat. Men det är inte alls säkert att det fungerar så i praktiken. I detta avsnitt undersöks om invandrare som tar ett personlighetstest i samband med ansökan om anställning svarar mer eller mindre uppriktigt på testet än övriga. De som svarar minst uppriktigt tenderar ju att få ett ”bättre” resultat på ett personlighetstest av självrapporttyp, och det är sådana test som dominerar i marknaden. I ”skarpt läge” är det mycket vanligt att de som testas ger en överdrivet positiv bild av sin personlighet. Alla skönmålar emellertid inte lika mycket. I förra avsnittet visade det sig att kvinnor som sökte chefsbefattningar svarade mera uppriktigt än vad män gjorde, vilket i motsvarande mån missgynnade dem. Frågeställningen är nu om samma problem finns vid testning av invandrare.

Metod

Data för denna rapport har insamlats i skarpt läge av rekryteringsfirman Katapult, som använder screeningmodulen av UPP-testet [109]. Detta är en förkortad version av testet som mäter 5 personlighetsdimensioner samt tendensen till skönmålning. Se vidare avsnitt 19. Alla personlighetsvariablerna korrigeras för skönmålning med hjälp av en psykometrisk metodik som beskrivs i avsnitt 7 och i [105]. Denna metodik har validerats både i experimentell forskning och med hjälp av analyser av testdata från arbetsökande.

Av de testade var 1372 av svensk härkomst, 250 utländska. De fördelade sig som sökande till ett 50-tal olika jobb. Ålder registrerades inte i datafilen. Sextiofyra procent var kvinnor.

Resultat

I Tabell 14.2 ges reliabilitetsdata för de fem personlighetsvariablerna samt måttet på skönmålning.

Tabell 14.2. Reliabiliteter (alfa-värden) för personlighetsvariablerna, samt korrelation med tendensen till skönmålning.

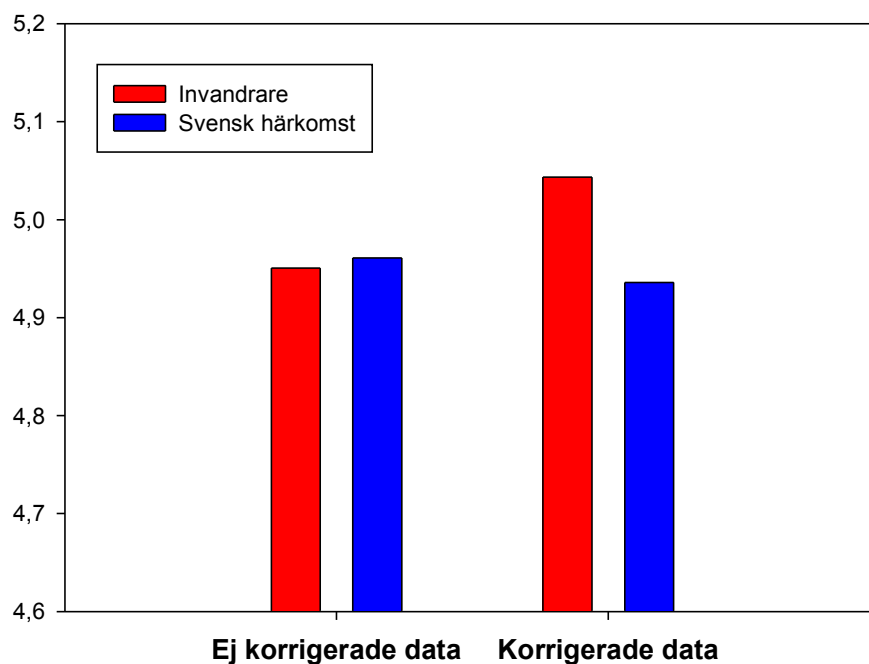
Skönmålning	Svensk härkomst		Utländsk härkomst	
	Reliabilitet	Korrelation med tendensen till skönmålning	Reliabilitet	Korrelation med tendensen till skönmålning
Utåtvändhet	0.74	0.35	0.82	0.39
Uthållighet	0.65	0.48	0.61	0.53
Samarbetsvilja	0.73	0.58	0.75	0.56
Positiv grundhållning	0.77	0.50	0.72	0.52
Kreativitet	0.67	0.18	0.63	0.14
Skönmålning	0.63	-	0.65	-

Tabellen visar att reliabiliteterna var acceptabla och på samma nivå i båda grupperna. Den visar också att relationen till skönmålning var av samma storleksordning i båda grupperna, medan den varierade mellan personlighetsvariablerna. Detta är ett vanligt resultat [109] och ett skäl till att korrektionen för skönmålning måste anpassas särskilt för varje personlighetsvariabel.

Alla skalor har standardiserats till medelvärde=5 och standardavvikelse=2. Medelvärdet i skönmålning var 4.77 för invandrare och 5.05 för dem med svensk härkomst. Skillnaden var statistiskt signifikant: $t=2.04$, $p=0.043$.

I figurerna 14.4 – 14.8 visas medelvärden i personlighetsvariablerna före och efter korrektion för skönmålning. Figur 14.4 ger resultaten för utåtvändhet.

Utåtvändhet

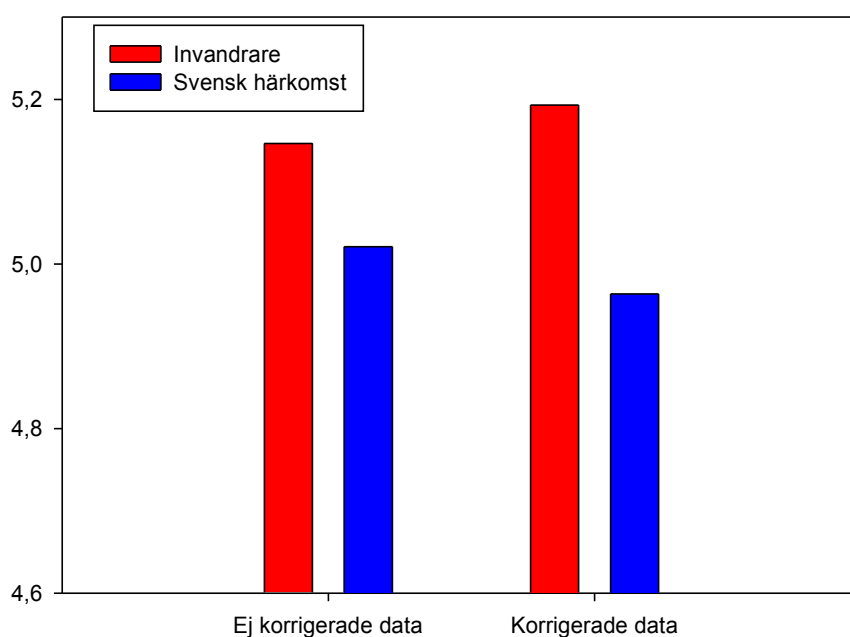


Figur 14.4. Medelvärden i utåtvändhet före och efter korrektion för skönmålning.

I okorrigerade data fanns ingen skillnad mellan grupperna i utåtvändhet medan invandrargruppens medelvärde var högst efter korrektion för skönmålning.

Figur 14.5 ger resultaten för kreativitet.

Kreativitet

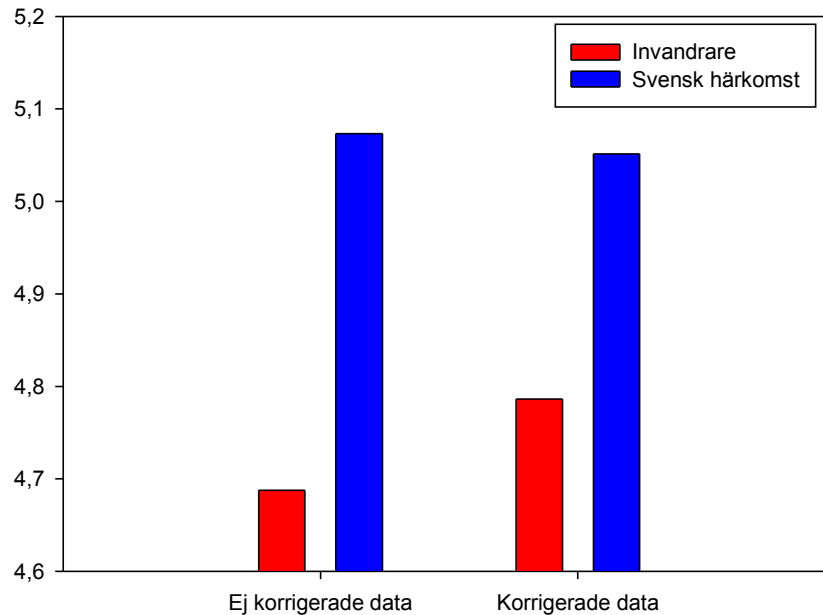


Figur 14.5. Medelvärden i kreativitet före och efter korrektion för skönmålning.

Figuren visar ett högre värde i kreativitet för invandrarna än för övriga, en skillnad som blev större efter korrektion för skönmålning.

Resultaten för samarbetsvilja ges i figur 14.6.

Samarbetsvilja

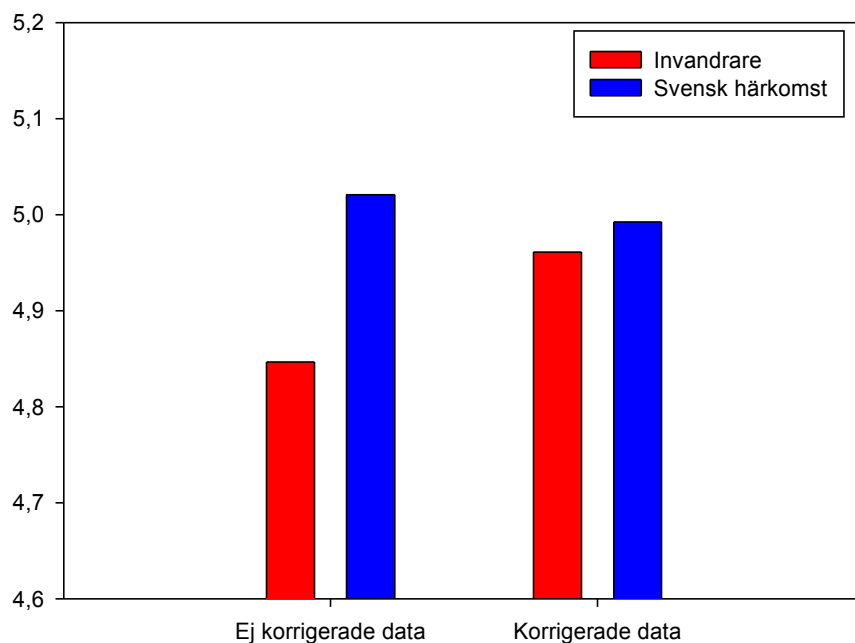


Figur 14.6. Medelvärden i samarbetsvilja före och efter korrektion för skönmålning.

Figuren visar ett lägre värde i samarbetsvilja för invandrarna men skillnaden mellan grupperna minskade betydligt efter korrektion för skönmålning.

Figur 14.7 ger resultaten för uthållighet.

Uthållighet

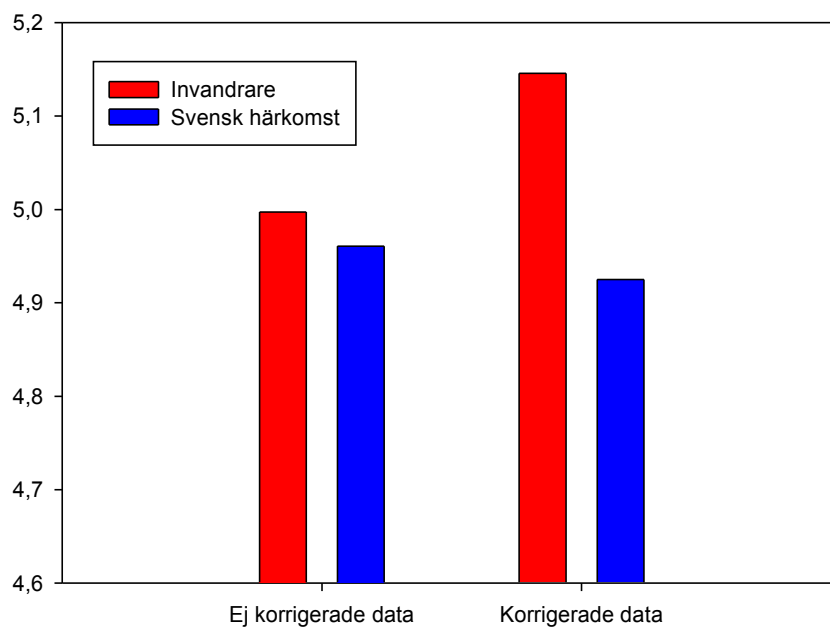


Figur 14.7. Medelvärden i uthållighet före och efter korrektion för skönmålning.

Figuren visar en nackdel för invandrarna med avseende på uthållighet i okorrigerade data men den eliminerades efter korrektion för skönmålning.

Resultaten för positiv attityd ges i figur 14.8.

Positiv attityd



Figur 14.8. Medelvärden i positiv hållning före och efter korrektion för skönmålning.

Invandrarna uppvisade en mera positiv hållning. Tendensen var speciellt tydlig efter korrektion för skönmålning.

Skillnaderna mellan okorrigerade och korrigerade värden var tydliga och trenden genomgående densamma: invandrarna gynnades av korrektionen för skönmålning. Effekterna var emellertid små när de studerades på medelvärdesnivå. Frågan är hur mycket de betydde för rangordningen av enskilda individer.

För att besvara den frågan rangordnades varje personlighetsvariabel före och efter korrektion för skönmålning. Varje persons rangordningstal bestämdes och skillnaden beräknades mellan hans eller hennes rangtal före och efter korrektionen. En positiv skillnad betyder att personen fick ett högre värde i rangordningen medan en negativ skillnad betydde att han eller hon förlorade i rangordningen. Tabell 14.3 visar de genomsnittliga rangförändringarna.

Personlighetsvariabel	Svensk härkomst	Invandrare
Utåtvändhet	-4.6	29.3
Uthållighet	-5.1	27.7
Samarbetsvilja	-5.5	39.3
Positiv grundhållning	-6.5	35.4
Kreativitet	-1.4	9.1
Medelvärde	-4.6	28.2

Även i dessa analyser är trenden tydlig och genomgående densamma. Att skillnaden i rangplacering är relativt liten i kreativitetsvariabeln beror troligen på att sambandet mellan den variabeln och skönmålning var ganska svagt.

Man kan diskutera hur viktiga dessa effekter är. För den som söker ett arbete, eller är angelägen om mångfald, är emellertid en förändring i rangplacering med ca 30 steg knappast en ointressant effekt.

Ännu en analys tar speciell hänsyn till extrema effekter av korrektionen för skönmålning. Eftersom det var många sökande till varje jobb²⁴ är extremgruppen särskilt intressant. De 2 % som visade den största förbättringen i sin genomsnittliga rangplacering efter korrektion studerades särskilt. Chansen till en rangförändring bland de 2 % mest fördelaktiga förändringarna var dubbelt så stor för invandrarna (3.2 %) som för övriga (1.6 %). Detta innebär troligen en kraftig förbättring av chanserna att få det sökta jobbet, om testresultatet spelar stor roll i processen.

Slutsats om mångfald

I detta avsnitt har resultat, baserade på ett stort datamaterial insamlat i skarpt läge från jobbsökande, presenterats som visar att invandrare skönmålade mindre än sökande med svensk härkomst. Samtidigt visade resultaten att testet (screeningmodulen av UPP-testet) hade lika goda psykometriska egenskaper i båda grupperna. Invandrarna missgynnades av en testutvärdering som inte tog hänsyn till skönmålning. Särskilt för den lilla grupp som hade

²⁴ Ca 40 sökande/jobb, i genomsnitt.

chans att få ett jobb (det var ca 40 sökande till varje jobb i genomsnitt), var korrektionen viktig för att ge invandrarna en rättvis chans att komma ifråga.

UPP-testet verkar alltså för rättvisa både mellan män och kvinnor och mellan invandrare och personer med svensk härkomst.

15. Normativt eller ipsativt svarsformat?

UPP-testet använder ett normativt svarsformat, dvs. varje testuppgift besvaras på en skala med fem steg. En sådan skala brukar kallas Likert-skala och används mycket ofta både för attitydmätning och i personlighetstestning. Problemet är att somliga testade personer skönmålar. *UPP*-testet använder en effektiv metod för att rensa bort effekterna av skönmålning.

En vanlig metodik som har samma syfte är att använda ett ipsativt svarsformat. Idén är att man ber den testade personen att jämföra vilket av två eller flera svarsalternativ som är bäst. Om alla alternativen är ungefär lika socialt önskvärda skulle man på det sättet kunna korrigera för tendensen till skönmålning. I den tekniska rapporten, Bilaga 3, diskuteras den psykometriska litteraturen om ipsativa och normativa svarsformat. Från den diskussionen kan vissa slutsatser dras:

- Ipsativa format leder till stora statistiska och psykometriska problem
- Ipsativa format lyckas inte med att korrigera bort effekterna av skönmålning
- Ipsativa format är impopulära bland dem som testas eftersom det känns orimligt och absurt att ”välja mellan äpplen och päron”.²⁵

I den tekniska rapporten förs frågorna ett steg vidare (Avsnitt 13). Den normativa informationen från testning med *UPP*-testet överfördes i ipsativ form genom att bara variationen inom individer togs tillvara, inte variationen mellan dem. I valideringsanalyser visade det sig sedan att det normativa formatet var överlägset och att det ipsativa inte bidrog utöver det normativa.

Alltså: det normativa formatet i *UPP*-testet fungerar bra; problemet med skönmålning kan lösas. Ingenting talar för att ett ipsativ format skulle fungera bättre. Tvärtom finns det starka argument för motsatsen. Som påpekats i avsnitt 9 finns det andra skäl för en ipsativ analys och en sådan görs också av *UPP*-testet.

16. Jasägartendensen i betydelsen globalt jasägande

I ett arbete av Couch och Keniston [22] har frågan om en jasägartendens diskuterats. Denna skulle innebära att testade personer varierar i vilket avseende de håller med om påståenden, oavsett deras innebörd. Det *finns* faktiskt en sådan tendens, men den är svag. I den tekniska rapporten (Avsnitt 13) utreds saken närmare.

Om alla testuppgifter i *UPP*-testet betraktas som mått på en och samma skala oavsett innebörd får vi en hög alfa-koefficient = 0,83. Detta värde uppkom trots att den

²⁵ Frågor av den här typen kan förekomma: ”Sitter du helst hemma och läser en god bok *eller* Har du lätt att komma på nya idéer?”

genomsnittliga interkorrelationen endast var = 0,02. Skälet till det höga alfa-värdet var att så många som 201 testuppgifter ingick i skalan på ja-sägartendens.

Har denna tendens någon praktisk betydelse? Man brukar korrigera för den genom att ha lika många uppgifter i båda riktningarna i varje skala. I en skala som mäter emotionell stabilitet ska man, för att ta ett exempel, ha lika många uppgifter som mäter instabilitet som stabilitet. Men eftersom jasägartendens är så svag är det inte säkert att kravet är viktigt. I *UPP*-testet är det flera skalor som inte är strikt balanserade i enlighet med det nämnda kravet. Det är inte alltid lätt att konstruera uppgifter i båda riktningarna, t ex för att sådant begrepp som perfektionism. Vad är motsatsen till perfektionism? Slarvighet eller normal ordningsamhet?

I den tekniska rapporten visas att jasägartendensens samvariation med delskalorna i *UPP*-testet var svag och saknade tydligt samband med hur balanserade skalorna var. Slutsatsen är att testresultaten inte är påverkade av jasägartendens. Den existerar men är svag och saknar praktisk betydelse, åtminstone för *UPP*-testet.

Det är viktigt att skilja denna traditionella definition av jasägartendens från den som används vid mätning av testdatas kvalitet. Den senare jasägartendens grundas på tendensen att hålla med om positivt formulerade uppgifter *och* att avvisa negativt formulerade uppgifter. Se avsnitt 8.

17. Varför bättre än traditionella test?

Det finns tre sannolika skäl till att *UPP*-testet är bättre än ett traditionellt Big Five-test.

1. Den traditionella FFM-modellen är kognitivt orienterad och den testade rapporterar om sig själv i ett utifrånperspektiv. *UPP*-testet lägger till viktiga dimensioner som mäter emotioner och gör det från ett inifrånperspektiv. ”En sådan person är jag” är den traditionella ansatsen. ”Så här känner jag mig” är *UPP*-testet. Det handlar om struktur kontra dynamik, om man så vill. Omfattande forskning har visat hur viktigt det dynamiska och emotionella perspektivet är, inte minst i arbetslivet. FFM tenderar att fastna i de semantiska nyanserna i det språk vi använder för att beskriva oss själva och andra. Det har ett visst intresse, men räcker inte särskilt långt, som resultaten i denna manual har visat. Ett slående exempel är att dimensionerna i FFM-modellen inte har något samband med arbetsvilja.
2. FFM-dimensionerna är uttryck för variation mellan människor som är abstrakt och övergripande. De personlighetsvariabler som ger viktiga tillskott till förklaringen av arbetsrelaterade dimensioner är emellertid betydligt smalare och mera fokuserade. Jämför t ex vänlighet med samarbetsvilja (omvändningen av passiv aggression). Tabell 12.8 visar att vänlighet ger ett ytterst litet tillskott till förklaring och förståelse av arbetsmotivation mm, medan samarbetsvilja dominerar i tabellen och är den bästa enskilda förklarande variabeln. Samarbetsvilja mäter i sin omvändning social dysfunktion i arbetet, och inte som vänlighet en mera allmän social attityd. De smalare och mera fokuserade variablernas överlägsenhet som så tydligt kommit fram i de studier som redovisas här, är ett specialfall av överlägsenheten hos proximala variabler

jämförda med distala sådana. Det finns många exempel på hur detta fungerar, också från ett närliggande område som attitydmätning [94-95].

3. Korrektionen för skönmålning gör att de testade troligen inte kan ge en starkt felaktig bild av sig själva, även om de försöker. Därigenom uppnås troligen ännu bättre validitet, och rättvisa. Det går inte att bluffa sig till ett jobb eller en utbildning, och personer som är ärliga (många kvinnor) missgynnas inte längre. Observera dock att korrektionen måste göras på ett systematiskt, kraftfullt och empiriskt beprövat sätt, som i *UPP*-testet, och utvärderas i förhållande till en stabil och omfattande normdatabas

18. Screeningmodulen²⁶

Ett screeningtest skall vara användbart för en snabb sortering av många sökande till en befattning. Genom att använda ett screeningtest kan rekryteringen snabbas upp och göras billigare; samtidigt tyder psykometrisk analys på att goda resultat kan förväntas, särskilt om man har många sökande (liten urvalskvot). *UPP*-testet har en särskild modul avsedd för screening. Den kan kompletteras med ett begåvningsstest, vilket har skett i de analyser som presenteras här. Även begåvningsstestet har utvecklats av Psykologisk Metod.

Screeningmodulen består av 59 uppgifter ur *UPP*-testet som mäter följande skalor:

- Uthållighet
- Samarbetsvilja
- Kreativitet
- Utåtvändhet
- Positiv grundattityd

Endast en av skolorna ingår i FFM-modellen (utåtvändhet). Dessutom ingick en av skalorna för mätning av skönmålning, den koverta. Samtliga skalor i detta avsnitt har korrigerats för skönmålning om inget annat sägs.

Urvalet av testdimensioner grundades på resultaten av proxyvalideringar som redovisats ovan i avsnitt 10. Modulen har integrerats i Katapults rekryteringsplattform och där besvarats av 3486 jobbsökande²⁷, som alltså samtliga har besvarat testet i skarpt läge. Ca 15 minuter krävs för att besvara uppgifterna i screeningmodulen.

Reliabilitet

Reliabiliteter för skalorna framgår av Tabell 18.1, som också innehåller information om hur stor del av variansen i varje skala som förklaras av skönmålning.

²⁶ Ett tack till Hampus Breistein och Katapult AB som tillhandahållit de data som analyseras här.

²⁷ April 2010, nya data tillförs kontinuerligt till databasen och förstärker normdata. Av integritetsskäl insamlas ej data om kön och ålder, däremot om utbildning.

Tabell 18.1. Psykometriska egenskaper hos om skalorna i screeningmodulen. N=3486.

Skala	Reliabilitet (Cronbachs alfa)	Antal uppgifter	Genomsnittlig korrelation mellan uppgifterna i skalan	Andel av variansen som förklaras av skönmålning
Uthållighet	0,65	8	0,19	0,183
Samarbetsvilja	0,83	11	0,30	0,409
Kreativitet	0,68	8	0,20	0,060
Utåtvändhet	0,81	10	0,29	0,326
Positiv grundattityd	0,76	10	0,25	0,354
Kovert skönmålning	0,69	12	0,16	-

Tabellen visar att skalorna hade tillfredsställande reliabilitet.

Skönmålning

Personlighetsskalorna var, i de flesta fallen, starkt korrelerade med skönmålning. Undantag fanns emellertid. Slutsatsen blir att skönmålning förekommer och är viktig, men att dess betydelse varierar mellan olika skalor och att ett globalt mått på skönmålning inte kan användas för att ta ställning till giltigheten i alla skalor.

Skönmålningen hos jobbsökande var mycket högre än i normdata (oskarpt läge). Skillnaden i medelvärdena (standardiserade skalor) var 1,06 – högt signifikant.

Medelvärdena före och efter korrektion för skönmålning, i oskarpt läge och bland jobbsökande, framgår av Tabell 18.2.

Tabell 18.2. Medelvärden (standardiserade skalor) i personlighetsvariablerna vid oskarpt läge och bland jobbsökande, för och efter korrektion för skönmålning med kovert skala.

Skala	Före korrektion		Efter korrektion	
	Oskarpt läge	Jobbsökande	Oskarpt läge	Jobbsökande
Uthållighet	-0,67	0,19	-0,20	0,05
Samarbetsvilja	-0,52	0,15	-0,03	0,01
Kreativitet	-0,43	0,12	-0,18	0,05
Utåtvändhet	-0,56	0,16	-0,18	0,05
Positiv grundattityd	-0,63	0,17	-0,13	0,04

Tabellen visar att en mycket stor del av effekten av skönmålning eliminerades, i genomsnitt 74 %.

Testdatas kvalitet

Testdatas kvalitet har utvärderats med tre mått:

- Intra-individuell spridning – i vilken mån den testade differentierar sina svar mellan olika testuppgifter
- Jasägandetendens – i vilken mån den testade tenderar att bejaka positiva påståenden och förneka negativa påståenden²⁸
- Strukturlikhet – i vilken mån svaren som en person ger på testuppgifterna samvarierar med gruppens medelvärden²⁹

Tabell 18.3 ger korrelationerna mellan dessa tre mått.

Kvalitetsskala	Intra-individuell spridning	Jasägandetendens	Strukturlikhet
Intra-individuell spridning	1,00	0,91	0,62
Jasägandetendens	0,91	1,00	0,83
Strukturlikhet	0,62	0,83	1,00

Som synes var sambanden starka. De tre måtten kunde kombineras till en sammansatt kvalitetsskala, vars reliabilitet (Cronbachs alfa) var = 0,87. Detta sammanslagna mått på kvalitet hade tydliga samband med alla personlighetsskalorna, se Tabell 18.4

Skala	Korrelation med datakvalitet
Uthållighet	0,31
Samarbetsvilja	0,32
Kreativitet	0,29
Utåtvändhet	0,42
Positiv grundattityd	0,38

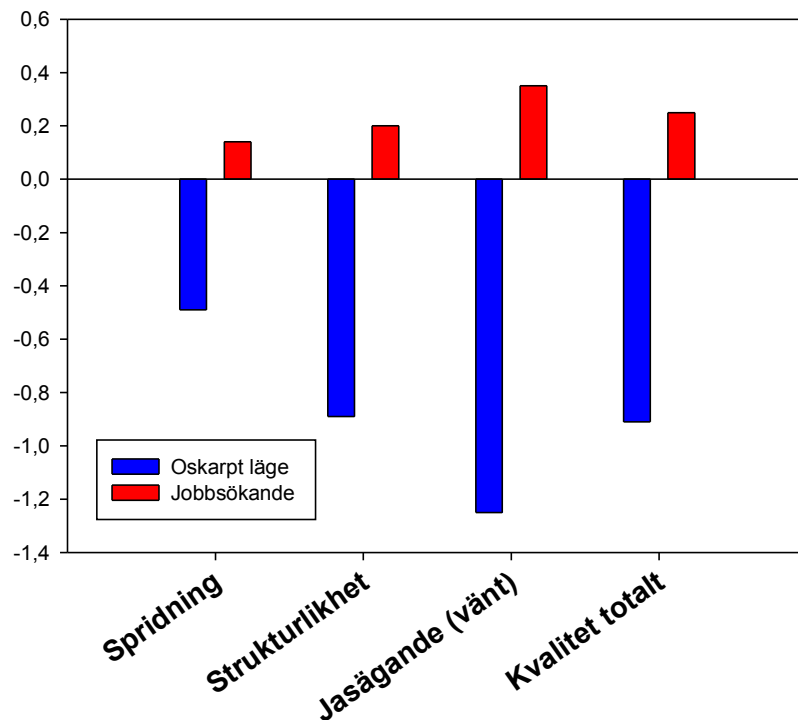
Samtliga värden var statistiskt signifikanta.

Datakvaliteten var dramatiskt högre i data från jobbsökande insamlade med screeningmodulen än data insamlade för normdatabasen, i oskarpt läge. Se Fig. 18.1.

²⁸ Observera att positivt och negativt orienterade påståenden behandlas olika. Jasägande på alla uppgifter oavsett orientering är en mycket svag faktor som inte behandlas här.

²⁹ Tanken bakom detta mått är att den som svarar vårdslöst eller inte förstår uppgifterna får en lägre sådan korrelation.

Medelvärde, standardiserad skala



Figur 18.1 Genomsnitt av mått på testdatas kvalitet i skarpt och oskarpt läge.

Utbildningsnivå

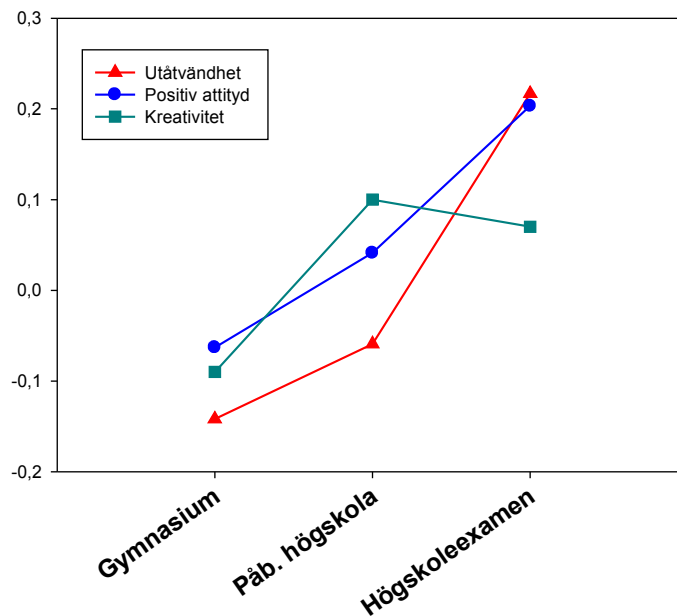
Två av personlighetsskalorna hade tydliga samband med utbildningsnivå, se Tabell 18.5 och Fig. 18.2

Tabell 18.5 Korrelationer mellan testskalorna och utbildningsnivå .	
Skala	Korrelation med utbildningsnivå
Uthållighet	-0,02
Samarbetsvilja	0,10**
Kreativitet	0,07**
Utåtvändhet	0,13**
Positiv grundattityd	0,04
Kovert skönmålning	0,01

** $p < 0,01$

Sambanden var som synes svaga även om tre av dem var signifikanta. Medelvärden för dessa tre personlighetsvariabler har i Fig. 18.2 prickats mot utbildningsnivå.

Medelvärde, standardiserad skala



Figur 18.2. Samband mellan 3 personlighetsvariabler och uppnådd utbildningsnivå, screeningmodulen.

Begåvningsmätning

Den version av screeningmodulen som användes innehöll också ett test som mäter begåvning, ett analogitest kallat PM2, med 19 uppgifter³⁰. Reliabiliteten var 0,74, genomsnittlig korrelation mellan uppgifterna 0,13³¹. Korrelationer mellan begåvningsstestet och personlighetsskalorna samt utbildningsnivå framgår av Tabell 18.6.

Tabell 18.6 Korrelationer mellan personlighetsskalorna, utbildningsnivå och begåvningsstestet.

Skala	Korrelation med begåvningsstestet
Uthållighet	0,00
Samarbetsvilja	0,20**
Kreativitet	0,08**
Utåtvändhet	0,12**
Positiv grundattityd	0,05**
Kovert skönmålning	0,18**
Utbildningsnivå	0,30**

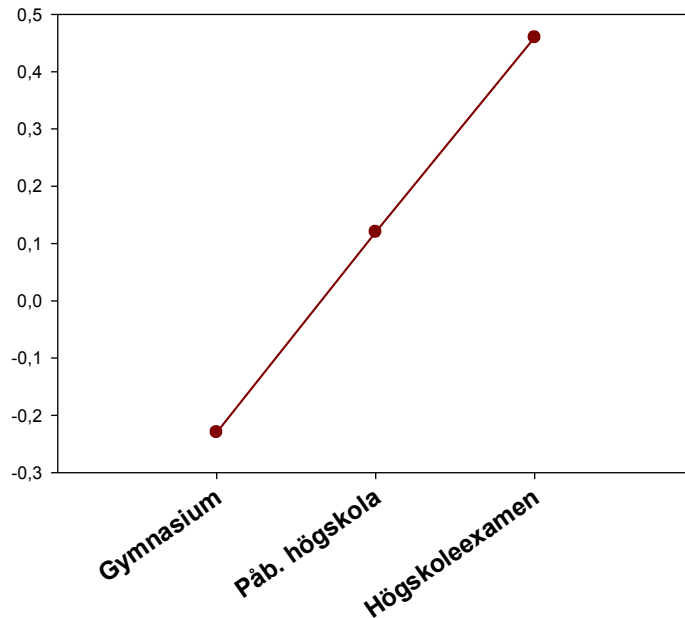
$p < 0,01$

³⁰ En av de ursprungligen 20 uppgifterna utgick på grund av låg reliabilitet.

³¹ Detta test togs i samband med den övriga Internettestningen. En viss risk kunde finnas att uppgifterna blev kända och att de testade därför förberedde sig. Korrelationen mellan testresultat och tidpunkt för testningen var emellertid endast = 0.08. Tidsperioden var ganska lång: sommaren 2009 – april 2010, så risken för spridning och träning på uppgifterna tycks vara obefintlig. Sedan april 2010 använder Katapult ett annat test eftersom man ville ha ett icke-verbalt test. Skillnaden mellan invandrare och övriga på begåvningsstestet var i genomsnitt $d=0,4$, till de senares nackdel, vilket ju talar för att det är lämpligt att gå över till ett icke-verbalt test om många som testas inte har svenska som modersmål. Internetbaserad testning av personlighet och begåvning tycks inte ge distorderade resultat [1].

Tabellen visar på signifikanta men låga samband mellan personlighetsskalorna och begåvning, vilket stämmer bra med resultat för hela UPP-testet som redovisats i Avsnitt 12. Sambandet mellan testresultat och begåvningstestet illustreras också i Fig. 18.3.

Medelvärde i begåvning, standardiserad skala



Figur 18. 3. Medelvärden i begåvningstestet för olika utbildningsnivåer.

Validering

Screeningmodulens variabler har proxyvaliderats med hjälp av data från testningar med hela UPP-testet, som tagits av 971 (oskarpt läge), resp. 144 personer (skarpt läge). Se Tabell 18.7.

Tabell 18.7 Korrelationer mellan personlighetsskalorna och genomsnittet av sex proxy-kriterier

Skala	Skarpt läge	Övriga
Uthållighet	0,32	0,26
Samarbetsvilja	0,23	0,26
Kreativitet	0,21	0,25
Utåtvändhet	0,31	0,29
Positiv grundattityd	0,72	0,73

Tabellen visar på god validitet särskilt för skalan Positiv grundattityd, men även de övriga skalorna har tydliga samband med proxykriterierna. Alla korrelationerna i tabellen är signifikanta. Skarpt läge ger ungefär samma resultat som för övriga testade, i detta avseende.

Slutsatser

Sammanfattningsvis kan konstateras att

- Screeningmodulens fem skalor hade god reliabilitet och validitet i betydelsen proxyvaliditet
- Skönmålningen var som väntat dramatiskt högre bland jobbsökande än bland dem som tog testet i oskarpt läge
- Korrektionen för skönmålning eliminerade i genomsnitt 74 % av dess effekt
- Mätningarna av testdatas kvalitet hade goda egenskaper; kvalitetsmåten samvarierade positivt med de fem personlighetsskalorna
- Testdatas kvalitet var betydligt högre hos de jobbsökande än i normmaterialet (bortsett från skönmålningen)
- Tre av de fem skalorna hade signifikanta samband med utbildningsnivå
- Det använda begåvnings-testet hade god reliabilitet och högt samband med utbildningsnivå, men låga samband med personlighetsskalorna – resultat som validerar både begåvnings-testet och personlighetsskalorna.

Resultaten torde vara mycket reliabla då de bygger på en mycket stor datamängd. Denna datamängd fungerar nu som normgrupp både för personlighetsskalorna och för begåvnings-testet vid användning av screening-testet och screeningmodulen³². Testdatas kvalitet kan också utvärderas gentemot denna datamängd för varje testad person. Låg kvalitet är ett varningstecken och kan tyda på att testningen bör göras om eller ges mindre vikt.

Eftersom både personlighet och begåvning är viktiga prognosfaktorer i arbetslivet kan screening-testet vara till nytta genom att den erbjuder er snabb första utvärdering i situationer där man har många sökande till en befattning eller utbildning. I sådant praktiskt arbete är det av avgörande betydelse att man kan korrigera för skönmålning, som är mycket vanlig och stark vid testning i skarpt läge, och att det är möjligt att få en mätning av testdatas kvalitet som är reliabel och valid. Urvalet av personlighetsvariabler för screeningmodulen gjordes utifrån valideringsstudier som visade att fokuserade variabler fungerar bättre än de traditionella FFM-dimensionerna (med ett undantag).

³² Modulen innehåller endast personlighetsskalor, ingen begåvningsmätning.

19. Källor till fördjupade kunskaper

Om *UPP*-testet – teori, utveckling, utvärdering:

http://swoba.hhs.se/hastba/abs/hastba2008_007.htm

Om personlighetstest i allmänhet:

http://swoba.hhs.se/hastba/abs/hastba2000_009.htm

Mabon, H. (2004) *Arbetspsykologisk testning. Om urvalsmetoder i arbetslivet. Andra upplagan*. Psykologiförlaget, Stockholm.

Om testmetodik:

Magnusson, D. (2003). *Testteori*. Stockholm: Psykologiförlaget.

Fhanér, S. (1986/1968) *Psykologisk testning : teori och tillämpningar*. Stockholm: Norstedt.

Hemsida för Psykologisk Metod AB:

www.psykologisk-metod.se

Professor Lennart Sjöbergs forskning:

<http://www.dynam-it.com/lennart/>

20. Referenser

- [1]. Arthur, W., Jr., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored Internet-based tests of cognitive ability and personality. [doi:10.1111/j.1468-2389.2010.00476.x]. *International Journal of Selection and Assessment*, 18(1), 1-16.
- [2]. Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology*, 49, 141-168.
- [3]. Bandura, A. (1986). *Social foundations of thought and action*. Englewood-Cliffs, NJ: Prentice-Hall.
- [4]. Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology*, 88(1), 87-99.
- [5]. Bateman, T. S., & Crant, J. M. (1993). The proactive component of organizational behavior: A measure and correlates. *Journal of Organizational Behavior*, 14(2), 103-118.
- [6]. Bettencourt, L. A., Gwinner, K. P., & Meuter, M. L. (2001). A comparison of attitude, personality, and knowledge predictors of service-oriented organizational citizenship behaviors. [doi:10.1037/0021-9010.86.1.29]. *Journal of Applied Psychology*, 86(1), 29-41.
- [7]. Björklund, C. (2001). *Work motivation - Studies of its determinants and outcomes*. Stockholm School of Economics, Stockholm.
- [8]. Björklund, C. (2007). Work motivation and perceived risks. *International Journal of Risk Assessment and Management*, 7(2), 237-247.
- [9]. Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, 10, 99-109.
- [10]. Bowman, M. L. (2002). The Misinterpretation of Percentiles. *Archives of Clinical Neuropsychology*, 17, 295-303.
- [11]. Brown, A., & Bartram, D. (2009). *Development and psychometric properties of the OPQ32r. Supplement to the OPQ 32 technical manual*: SHL.
- [12]. Brown, J. D., & Dutton, K. A. (1995). The thrill of victory, the complexity of defeat: Self-esteem and people's emotional reactions to success and failure. *Journal of Personality & Social Psychology*, 68(4), 712-722.
- [13]. Brown, S. P., Cron, W. L., & John W. Slocum, J. (1997). Effects of Goal-Directed Emotions on Salesperson Volitions, Behavior, and Performance: A Longitudinal Study. *Journal of Marketing*, 61(1), 39-50.
- [14]. Brown, T. J., Mowen, J. C., Donovan, D. T., & Licata, J. W. (2002). The customer orientation of service workers: Personality trait effects on self- and supervisor performance ratings. [doi:10.1509/jmkr.39.1.110.18928]. *Journal of Marketing Research*, 39(1), 110-119.
- [15]. Caprara, G. V., & Cervone, D. (2000). *Personality: Determinants, dynamics, and potentials*: Cambridge University Press: New York.
- [16]. Caputo, D., & Dunning, D. (2005). What you don't know: The role played by errors of omission in imperfect self-assessments. *Journal of Experimental Social Psychology*, 41(5), 488-505.
- [17]. Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. [doi:10.1207/s15327906mbr1203_2]. *Multivariate Behavioral Research*, 12(3), 289-325.
- [18]. Christiansen, N. D., Janovics, J. E., & Siers, B. P. (2010). Emotional intelligence in selection contexts: Measurement method, criterion-related validity, and vulnerability to

- response distortion. [doi:10.1111/j.1468-2389.2010.00491.x]. *International Journal of Selection and Assessment*, 18(1), 87-101.
- [19]. Cliff, N., & Keats, J. A. (2003). *Ordinal Measurement in the Behavioral Sciences*. Mahwah, NJ: Erlbaum.
- [20]. Cohen, J. (1988). *Statistical power analysis for behavioral sciences (2nd ed.)*. Hillsdale, NJ: Erlbaum.
- [21]. Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- [22]. Couch, A., & Keniston, K. (1961). Agreeing response set and social desirability. *The Journal of Abnormal and Social Psychology*, 62(1), 175-179.
- [23]. Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting and Clinical Psychology*, 24, 349-354.
- [24]. Ding, D. X., Hu, P. J.-H., Verma, R., & Wardell, D. G. (2010). The impact of service system design and flow experience on customer satisfaction in online financial services. [doi:10.1177/1094670509350674]. *Journal of Service Research*, 13(1), 96-110.
- [25]. Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087-1101.
- [26]. Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest*, 5(3), 69-106.
- [27]. Ekman, G. (1952). *Differentiell psykologi*. Stockholm: Almqvist & Wiksell.
- [28]. Engelberg, E., & Sjöberg, L. (2004). Internet use, social skills, and adjustment. *Cyberpsychology & Behavior*, 7(1), 41-47.
- [29]. 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.
- [30]. Ferris, G. R., Treadway, D. C., Kolodinsky, R. W., Hochwarter, W. A., Kacmar, C. J., Douglas, C., et al. (2005). Development and validation of the political skill inventory. *Journal Of Management*, 31(1), 126.
- [31]. Fhanér, S. (1986/1968). *Psykologisk testning : teori och tillämpningar*. Stockholm: Norstedt.
- [32]. Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research*, 14, 449-468.
- [33]. Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality & Individual Differences*, 7(3), 385-400.
- [34]. Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.
- [35]. Groth, M., & Goodwin, R. E. (2011). Customer service *APA handbook of industrial and organizational psychology, Vol 3: Maintaining, expanding, and contracting the organization*. (pp. 329-357): Washington, DC, US: American Psychological Association.
- [36]. Guilford, J. P. (1950). Creativity. *American Psychologist*, 5, 444-454.
- [37]. Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- [38]. Hayman, J. (2005). Psychometric assessment of an instrument designed to measure work life balance. *Research and Practice in Human Resource Management*, 13(1), 85-91.
- [39]. Heatherton, T. F., Wyland, C. L., & Lopez, S. J. (2003). Assessing self-esteem. In C. R. Snyder (Ed.), *Positive psychological assessment: A handbook of models and measures*. (pp. 219-233). Washington, (pp. 219-233). Washington, DC: American Psychological Association.
- [40]. Hedlund, N. (2008). Bilder och berättelser som kanske visar vårt inre. *Personal och Ledarskap*, 39(10), 20-27.

- [41]. Heneman, H. G. (1974). Comparisons of self- and superior ratings of managerial performance. *Journal of Applied Psychology*, 59(5), 638-642.
- [42]. Hoffmann, R. M. (1995). Silent rage: Passive-aggressive behavior in organizations. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 56, 1138.
- [43]. Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, 88(1), 100-112.
- [44]. Hogan, R., & Hogan, J. (1997). *Hogan Development Survey*.: Hogan Assessment Systems.
- [45]. Hogan, R., Roberts, B. W., Walsh, W. B., Craik, K. H., & Price, R. H. (2000). A socioanalytic perspective on person-environment interaction. In K. H. Craik & R. H. Price (Eds.), *Person-environment psychology: New directions and perspectives (2nd ed.)* (pp. 1-23). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- [46]. Hopwood, C. J., Morey, L. C., Markowitz, J. C., Pinto, A., Skodol, A. E., Gunderson, J. G., et al. (2009). The construct validity of passive-aggressive personality disorder. [doi:10.1521/psyc.2009.72.3.256]. *Psychiatry: Interpersonal and Biological Processes*, 72(3), 256-267.
- [47]. Hough, L. M. (1998). The millennium for personality psychology: New horizons or good ole daze. *Applied Psychology: An International Review*, 47(2), 233-261.
- [48]. Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35(1), 1-19.
- [49]. Johnson, J. W. (2004). Factors Affecting Relative Weights: The Influence of Sampling and Measurement Error. *Organizational Research Methods*, 7(3), 283-299.
- [50]. Johnson, J. W., & LeBreton, J. M. (2004). History and Use of Relative Importance Indices in Organizational Research. *Organizational Research Methods*, 7(3), 238-257.
- [51]. Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. [doi:10.1037/a0017286]. *Journal of Applied Psychology*, 95(1), 54-78.
- [52]. Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluation traits - self esteem, generalized self-efficacy, locus of control, and emotional stability - with job satisfaction and performance: A meta-analysis. *Journal of Applied Psychology*, 86(1), 80-92.
- [53]. Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87(4), 765-780.
- [54]. Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127(3), 376-407.
- [55]. Kantor, M. (2002). *Passive-aggression: A guide for the therapist, the patient and the victim*: Praeger Publishers/Greenwood Publishing Group: Westport.
- [56]. Kihlstrom, J. F., & Cantor, N. (2000). Social intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 380-395). Cambridge, UK: Cambridge University Press.
- [57]. Kline, T. J. B., & Sulsky, L. M. (2009). Measurement and assessment issues in performance appraisal. [doi:10.1037/a0015668]. *Canadian Psychology/Psychologie canadienne*, 50(3), 161-171.
- [58]. Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.
- [59]. Lindley, P., Bartram, D., & Kennedy, N. (2008). *EFPA review model for the description and evaluation of psychological tests. Test review form and notes for reviewers. Version 3.42*: European Federation of Psychological Associations.

- [60]. Lyubomirsky, S., King, L., & Diener, E. (2005). The Benefits of Frequent Positive Affect: Does Happiness Lead to Success? *Psychological Bulletin*, *131*(6), 803-855.
- [61]. Lönnqvist, J.-E., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J., & Verkasalo, M. (2007). Substance and style in socially desirable responding. *Journal of Personality*, *75*(2), 291-322.
- [62]. Maddi, S. R. (2006). Hardiness: The courage to grow from stresses. *The Journal of Positive Psychology*, *1*(3), 160-168.
- [63]. Magnusson, D. (2003). *Testteori*. Stockholm: Psykologiförlaget.
- [64]. 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.
- [65]. Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits*, 2nd ed. Cambridge: Cambridge University Press.
- [66]. Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology*, *59*, 507-536.
- [67]. Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion*, *3*(1), 97-105.
- [68]. McIluff, E., & Coghlan, D. (2000). Reflections: Understanding and contending with passive-aggressive behaviour in teams and organizations. *Journal of Managerial Psychology*, *15*(7-8), 716-732.
- [69]. 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.
- [70]. Millon, T., Millon, C., Davis, R., & Grossman, S. (1997). *Millon Clinical Multiaxial Inventory-III [Manual Second Edition]*. Bloomington, MN: Pearson Assessments.
- [71]. Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- [72]. Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, *60*(4), 1029-1049.
- [73]. Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*(3), 683-729.
- [74]. Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, *10*, 71-83.
- [75]. Murphy, T., & Oberlin, L. H. (2005). *Overcoming passive aggression*. New York: Marlowe & Company.
- [76]. Neilson, G. L., Pasternack, B. A., & Van Nuys, K. E. (2005). The PASSIVE-AGGRESSIVE Organization. *Harvard Business Review*(10), 82-92.
- [77]. Nordlund, C. (2005). Att de inte skäms! *Chef*(11), 18-22.
- [78]. 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.
- [79]. Oreg, S. (2003). Resistance to change: Developing an individual differences measure. *Journal of Applied Psychology*, *88*(4), 680-693.
- [80]. Oreg, S., Bayazit, M., Vakola, M., Arciniega, L., Armenakis, A., Barkauskiene, R., et al. (2008). Dispositional resistance to change: Measurement equivalence and the link to personal values across 17 nations. *Journal of Applied Psychology*, *93*(4), 935-944.
- [81]. 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.

- [82]. Paulhus, D. L., & Reid, D. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, 60, 307-317.
- [83]. Pérez, J. C., Petrides, K. V., & Furnham, A. (2005). Measuring trait emotional intelligence. In R. Schulze & R. D. Roberts (Eds.), *International Handbook of Emotional Intelligence*. Cambridge, MA: Hogrefe & Huber.
- [84]. Ray, J. J. (1975). A behavior inventory to measure achievement motivation. *Journal of Social Psychology*, 95(1), 135-136.
- [85]. Ray, J. J. (1979). A quick measure of achievement motivation: Validated in Australia and reliable in Britain and South Africa. *Australian Psychologist*, 14(3), 337-344.
- [86]. Rice, K. G., & Aldea, M. A. (2006). State Dependence and Trait Stability of Perfectionism: A Short-Term Longitudinal Study. *Journal of Counseling Psychology*, 53(2), 205-213.
- [87]. Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2, 313-345.
- [88]. Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- [89]. Russell, D. (1996). The UCLA loneliness scale (version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment*, 66, 20-40.
- [90]. Schuerger, J. M., Zarrella, K. L., & Hotz, A. S. (1989). Factors that influence the temporal stability of personality by questionnaire. [doi:10.1037/0022-3514.56.5.777]. *Journal of Personality and Social Psychology*, 56(5), 777-783.
- [91]. Scratchley, L. S., & Hakstian, A. R. (2000-2001). The measurement and prediction of managerial creativity. *Creativity Research Journal*, 13, 367-384.
- [92]. Scroggins, W. A., Thomas, S. L., & Morris, J. A. (2009). Psychological testing in personnel selection, part III: The resurgence of personality testing. *Public Personnel Management*, 38(1), 67-77.
- [93]. Simonton, D. K. (1984). *Genius, creativity, and leadership*. Cambridge: Cambridge University Press.
- [94]. Sjöberg, L. (1980). Similarity and correlation. In E.-D. Lantermann & H. Feger (Eds.), *Similarity and choice* (pp. 70-87). Bern: Huber.
- [95]. Sjöberg, L. (1982). Attitude-behavior correlation, social desirability and perceived diagnostic value. *British Journal of Social Psychology*, 21, 283-292.
- [96]. Sjöberg, L. (1989). Mood and expectation. In A. F. Bennett & K. M. McConkey (Eds.), *Cognition in individual and social contexts* (pp. 337-348). Amsterdam: Elsevier.
- [97]. Sjöberg, L. (2001). Emotional intelligence: A psychometric analysis. *European Psychologist*, 6, 79-95.
- [98]. Sjöberg, L. (2003). Distal factors in risk perception. *Journal of Risk Research*, 6(3), 187-211.
- [99]. Sjöberg, L. (2004). *Personlighetstest i arbetslivet: historik och aktuell forskning* (SSE/EFI Working Paper Series in Business Administration No. 2000:9). Stockholm: Handelshögskolan, Sektionen för ekonomisk psykologi.
- [100]. Sjöberg, L. (2005). *Ledarskap och personlighet: Testning av jagstyrka. (Leadership and personality: Testing ego-strength)* (SSE/EFI Working Paper Series in Business Administration No. 2005:7). Stockholm: Stockholm School of Economics.
- [101]. Sjöberg, L. (2005). Självskattningstest: Den som ljuger fräckast vinner? (Self report tests: will the liar win?). *Personal och Ledarskap*, 36(6), 22-25.
- [102]. Sjöberg, L. (2007). Kreativitet och forskning. Under utarbetande.

- [103]. 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.
- [104]. Sjöberg, L. (2008). Emotional intelligence and life adjustment. In J. C. Cassady & M. A. Eissa (Eds.), *Emotional Intelligence: Perspectives on Educational & Positive Psychology* (pp. 169-183). New York: Peter Lang Publishing.
- [105]. Sjöberg, L. (2009). *UPP-testet: Korrektion för skönmålning. (The UPP test: Correction for impression management)*. *Forskningsrapport 2009:3*. Stockholm: Psykologisk Metod AB.
- [106]. Sjöberg, L. (2009). *UPP-testet: Kriterierelaterad validitet. Forskningsrapport 2009:2*. Stockholm: Psykologisk Metod AB.
- [107]. Sjöberg, L. (2010). *Emotionell intelligens och social förmåga hos ungdomar. Forskningsrapport 2010:4*. Stockholm: Psykologisk Metod AB.
- [108]. Sjöberg, L. (2010). *Faktorstrukturen hos UPP-testet. Forskningsrapport 2010:5*. Stockholm: Psykologisk Metod AB.
- [109]. Sjöberg, L. (2010). *Manual för UPP-testet*. Stockholm: Psykologisk Metod AB.
- [110]. Sjöberg, L. (2010). *UPP-testet och kundservice: Kriteriestudie. Forskningsrapport 2010:6*. Stockholm: Psykologisk Metod AB.
- [111]. Sjöberg, L. (2010). *UPP/Screen: Ett screeningtest av personlighet och begåvning. Forskningsrapport 2010:7*. Stockholm: Psykologisk Metod AB.
- [112]. 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.
- [113]. Sjöberg, L., & Lind, F. (1994). *Arbetsmotivation i en krisekonomi: En studie av prognosfaktorer. (Work motivation in a crisis economy: A study of prognostic variables)* (Studier i ekonomisk psykologi No. 121): Institutionen för ekonomisk psykologi, Handelshögskolan i Stockholm.
- [114]. Sjöberg, L., Littorin, P., & Engelberg, E. (2005). Personality and emotional intelligence as factors in sales performance. *Scandinavian Journal of Organizational Theory and Practice*, 15(2), 21-37.
- [115]. Sjöberg, L., & Möller, K. (2009). *Sociala arbetsfunktioner och personlighet*. Stockholm: Psykologisk Metod AB.
- [116]. Sjöberg, L., Svensson, E., & Persson, L.-O. (1979). The measurement of mood. *Scandinavian Journal of Psychology*, 20(1), 1-18.
- [117]. Smith, S. A., Kass, S. J., Rotunda, R. J., & Schneider, S. K. (2006). If at first you don't succeed: Effects of failure on general and task-specific self-efficacy and performance. *North American Journal of Psychology*, 8(1), 171-182.
- [118]. Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*, 124(2), 240-261.
- [119]. Stanley, K. D., & Murphy, M. R. (1997). A comparison of general self-efficacy with self-esteem. *Genetic, Social, and General Psychology Monographs*, 123(1), 79-99.
- [120]. Tsai, J. L., Knutson, B., & Fung, H. H. (2006). Cultural Variation in Affect Valuation. *Journal of Personality and Social Psychology*, 90(2), 288-307.
- [121]. Wall, T. D., Michie, J., Patterson, M., Wood, S. J., Sheehan, M., Clegg, C. W., et al. (2004). On the Validity of Subjective Measures of Company Performance. *Personnel Psychology*, 57(1), 95-118.
- [122]. Weis, S., & Süß, H.-M. (2007). Reviving the search for social intelligence - A multi-trait-multimethod study of its structure and construct validity. *Personality and Individual Differences*, 42(1), 3-14.

- [123]. Weisberg, R. W. (2006). *Creativity. Understanding innovation in problem solving, science, invention, and the arts*. New York: Wiley.
- [124]. Wetzler, S. (1992). *Living with the passive-aggressive man*. New York: Simon & Schuster.
- [125]. Vodanovich, S. J. (2003). Psychometric measures of boredom: A review of the literature. *Journal of Psychology: Interdisciplinary and Applied*, 137(6), 569-595.
- [126]. Vrij, A. (2000). *Detecting lies and deceit*. New York: Wiley.

Bilaga 1. Några begrepp inom testteorin

I manualen används några grundläggande begrepp i testteorin och de ska förklaras i denna bilaga. För en utförlig introduktion, se Magnussons lärobok i testteori [63] eller Fhanérs mera kortfattade och lättlästa lärobok [31]. För en full förståelse krävs naturligtvis att man tränger in i arbeten av den typen, men några av grundbegreppen kan man få en uppfattning med hjälp av nedanstående förklaringar:

Ett test av den typ som avses här består av ett antal *testuppgifter*, ofta kallade ”items” även på svenska. Här används den svenska termen testuppgift, eller helt enkelt uppgift. Varje uppgift består normalt av ett påstående och testpersonen ombeds ta ställning till i vilken mån det passar in på honom eller henne. Exempel:

”Jag blir nervös när jag ska tala inför en stor grupp av människor”.

Detta är en testuppgift som eventuellt mäter (brist på) emotionell stabilitet. Andra testuppgifter som avser att mäta emotionell stabilitet kan vara formulerade i motsatt riktning:

”Jag har nästan alltid ett lugnt och gott humör”

Alla uppgifter i testet bedöms av den testade personen på en svarsskala. I *UPP*-testet handlar det om fem steg:

Stämmer absolut
Stämmer i viss mån
Tveksamt
Stämmer knappast
Stämmer absolut inte

Dessa fem svarsalternativ tilldelas talen 5, 4, 3, 2 och 1 för fortsatta beräkningar.

Poängen på en testskala är medelvärdet³³ av svaren på alla uppgifter som ingår i skalan, efter det att svarsskalorna vänts för uppgifter som ger ett negativt uttryck för begreppet ifråga. I exemplet måste man alltså vända svarsskalan för den förstnämnda uppgiften så att svaren tillordnas talen 1, 2, 3, 4, och 5.

Ganska ofta möter man personer med teknisk/matematisk bakgrund som ställer sig tvekande till att använda ”lika skalsteg” för svarsalternativen. Om detta finns en omfattande diskussion [19]. Här får det räcka med att säga att det finns mycket stor erfarenhet av tillvägagångssättet och att inget tyder på att det skulle medföra några stora nackdelar. Däremot har det stora fördelar: det är enkelt och gör det möjligt att använda kraftfulla psykometriska och statistiska metoder.

Reliabilitet är avsaknad av slumpmässiga mätfel i en skala. Högsta tänkbara värdet är 1 då testet är helt fritt från slumpfel. Det vanligaste sättet att uppskatta reliabiliteten är alfa-koefficienten som också kallas homogenitetskoefficient. Denna har används genomgående i

³³ Ibland används summan, men det är mindre lämpligt i personlighetstest eftersom det medför att en person som hoppar över, eller inte hinner fram till, en uppgift, av det skälet får en lägre poäng i den skala där uppgiften ingår. Summan är adekvat när det gäller begåvningsstest, inte för personlighetstest.

arbetet med *UPP*-testet. Alfa ökar ju mera testuppgifterna samvarierar, och ju flera uppgifter en skala har. Reliabilitet kan också mätas som *stabilitet*, vilket kan uppskattas med hjälp av korrelationen mellan två mätningar med testet och samma grupp av testpersoner – låt oss säga med sex veckors mellanrum.

Reliabilitetskoefficienten är det vanligaste måttet på tillförlitligheten hos ett test. För vissa användningar är det mera lämpligt att beräkna *felspridningen*, standardavvikelsen hos mätfelen. Ju större den är desto större är osäkerheten i uttalanden om en enskild persons testresultat.

Validitet är ett mått på att testskalan mäter det som är avsett. Det mest rättframma sättet att mäta validitet är genom att uppskatta korrelationen mellan skalan och ett *kriterium*, dvs. en variabel som är ett mått på någon form av nytta. Det kan handla om arbetsresultat, sjukfrånvaro eller arbetsmotivation. En del kriterier är ”objektivt” bestämbara eller grundar sig på andra människors (ofta chefers) bedömningar, andra är uttryck för den testades egen upplevelse, motivation eller attityder. I det senare fallet, av stor betydelse för arbetet med *UPP*-testet, måste problemet med skönmålning särskilt uppmärksammas, varom mera i denna manual i avsnitt 6.

Validiteten kan också uppskattas genom att relatera en testskala till en annan skala som uppges mäta samma begrepp och som kanske är etablerad i omfattande och oberoende forskning och erfarenhet. Detta brukar kallas *begreppsvalidering*. I denna manual ges åtskilliga exempel på begreppsvalidering.

Vid begreppsvalidering vill man veta om en testvariabel verkligen mäter det begrepp den utger sig för att mäta. Det är därför lämpligt att korrigera korrelationerna mellan testvariabeln och kriterierna för mätfel. Mätfelen har den effekten att de sänker alla korrelationer. I manualen rapporteras både korrelationer mellan rådata och korrelationer efter korrektion för slumpmässiga mätfel.

Begreppsvalidering kan även bygga på experiment eller på samband mellan en testskala och annan information om de testade personerna. I denna manual redovisas ett experiment som validerar den använda metodiken för att korrigera för skönmålning, och data på samband mellan skalor på social förmåga, emotionell intelligens, den upplevda ensamheten och balans mellan arbete och övrigt liv.

Slutligen skall begreppen *normering* och *normgrupp* förklaras. Det ligger i sakens natur att testdata inte kan tolkas i absolut mening. Vilken nivå svaren ligger på beror på uppgifternas formulering. Man måste därför ha en tillräckligt stor normgrupp för varje skala. En testad persons resultat tolkas genom att det sätts i relation till normgruppen. Tillhör han de 10 % som ligger högst, är värdet ungefär genomsnittligt eller ligger han eller hon en bit under genomsnittet? Normgruppen måste vara av betydande storlek och behovet att insamla en sådan är ett av skälen till att testkonstruktion är en dyrbar och tidsödande process.

Bilaga 2. Kompletterande information om begreppsvalideringar

Social förmåga och emotionell intelligens

I denna bilaga ges resultat från begreppsvalidering som bygger på en särskild delstudie (studie 7) med studenter, normdata, studie 13 och Rapport 2009:1.

Deltagarna i en delstudie som syftade till begreppsvalidering var studerande inom grundutbildningen vid Mälardalens högskola, N=149. Av dessa var 110 kvinnor och 31 män. Åtta uppgav ej kön. Åldern varierade mellan 19 och 47 år, median = 24 år. De 149 besvarade en enkät med frågor om ensamhet, balans mellan arbete (studier) och övrigt liv, samt uppgifter som gick ut på att bedöma sociala problempisoder [97]. Av de 149 var det 67 som dessutom besvarade testet i Internetformat.

Det fanns 3 mått på emotionell intelligens: ett självrapportmått och två prestationsmått. Sambanden mellan dem återges i Tabell 1. Alla värden på emotionell intelligens i det följande är korrigerade för overta och koverta taktisksvar.

Tabell 1. Korrelationer mellan tre mått på emotionell intelligens samt social förmåga

	Episoder ³⁴	Ansiktsuttryck	Självrapportskala EI
Episoder	1,00	0,61**	0,31*
Ansiktsuttryck	0,61**	1,00	0,22
Självrapportskala EI	0,31*	0,22	1,00
Social förmåga	0,08	0,13	0,52**

** p < 0,01 * p < 0,05

Tabellen visar en hög grad av konvergens för de två prestationsmåten, lägre för dessa och självrapportskalan. Dock kan man se en viss grad av överensstämmelse även här. Detta är en vanlig typ av resultat vid mätning av emotionell intelligens [66]. Sambandet mellan EI självrapport och social förmåga var starkt. Tabell 2 återger korrelationerna mellan EI-måten, social förmåga, ensamhet och balans.

Tabell 2. Korrelationer mellan måten på emotionell intelligens och kriterier

	Episoder	Ansiktsuttryck	Självrapportskala EI	Social förmåga
Ensamhet	-0,22	-0,08	-0,50**	-0,50**
Balans	0,05	0,02	0,18	0,09

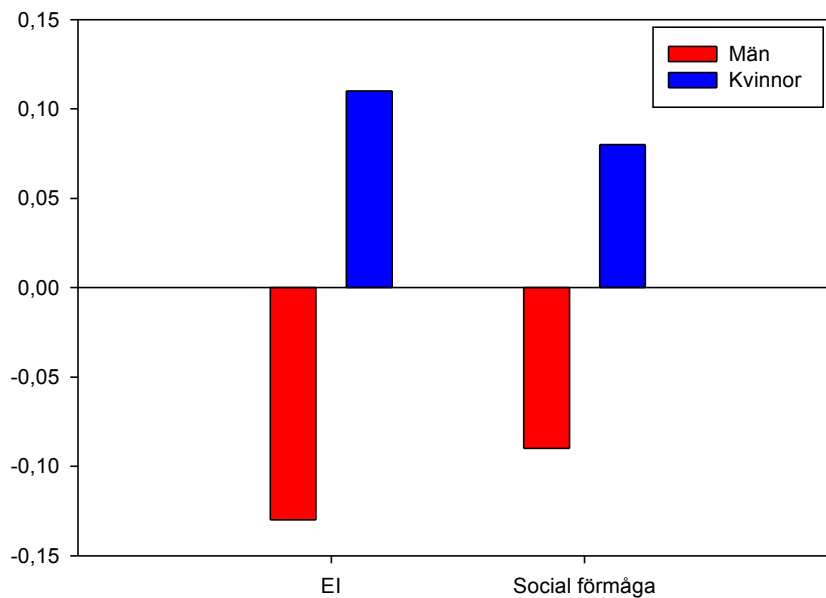
** p < 0.01

När det gäller ensamhet var sambanden som väntat, men däremot inte för balans. Ett samband mellan kön och emotionell intelligens förväntades emellertid. Kvinnor hade också betydligt högre värde i denna variabel än män. Skillnaden *t*-testades: $t(45) = 2,271, p = 0.028$. Samma

³⁴ Detta är ett prestationsmått på emotionell intelligens som beskrivs i den tekniska rapporten.

trend, med en stor skillnad, fanns för social förmåga men den var inte statistiskt signifikant. Ännu tydligare resultat erhöles i analys av normmaterialet, se Fig. 1. Här var även skillnaden i social förmåga statistiskt signifikant, även om den var mindre än skillnaden i emotionell intelligens.

Standardiserad testpoäng



Figur 1 . Genomsnitt i emotionell intelligens och social förmåga för män och kvinnor, standardiserade värden.

Civilstånd har i tidigare forskning befunnits ha ett samband med emotionell intelligens [104]. Ett samband är rimligt med tanke på att emotionell och social förmåga torde vara av betydelse för att upprätthålla en relation. När det gäller civilstånd var skillnaderna i denna undersökning avseende emotionell intelligens och social förmåga stora (absolut sett, inte i termer av statistisk signifikans). En envägs ANOVA gav $F(2,44) = 3,309, p = 0,046$. För social förmåga var motsvarande värde $F(2,44) = 2,410, p = 0,056$. Se Fig.2

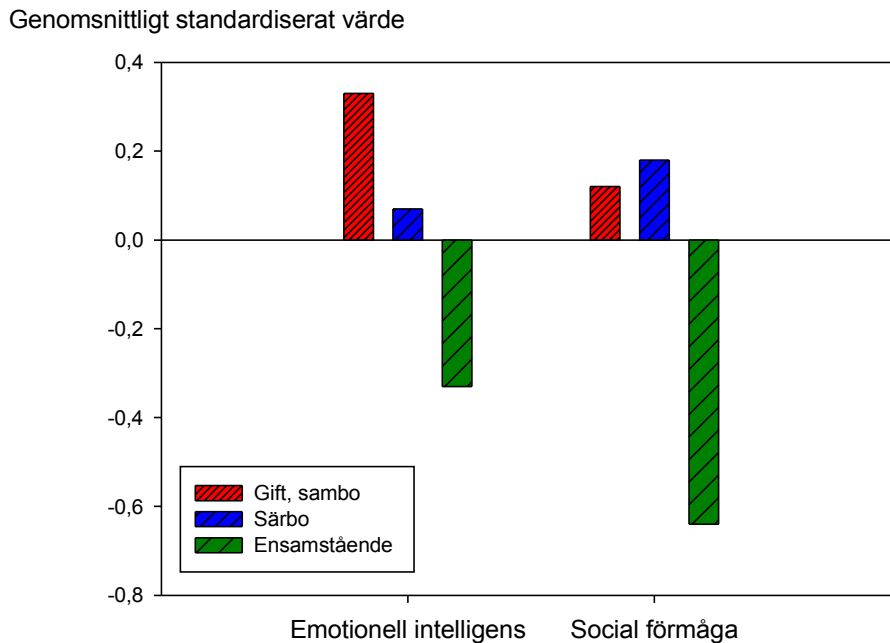
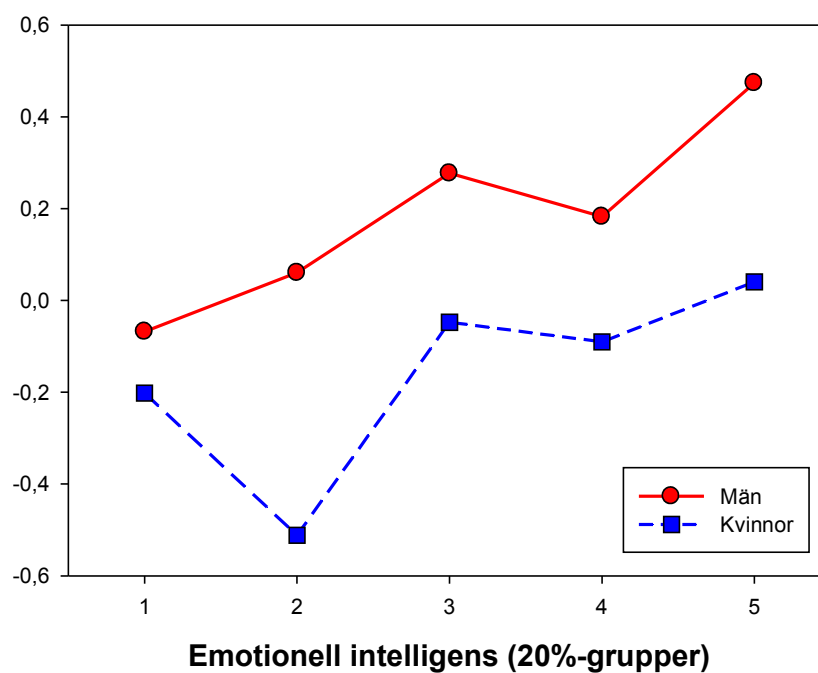


Fig. 2. Genomsnitt i emotionell intelligens och social förmåga i relation till civilstånd, standardiserade värden.

Problemet med balansvariabeln kunde lokaliseras till en könsskillnad som gick på tvären mot skillnaden i EI: kvinnorna hade lägre balans. Medelvärdena för män och kvinnor var 0,40 och -0,11, $t = 2,470$, $p = 0,015$. Vid en analys av sambandet mellan EI och balans för könsggrupperna separat var korrelationerna, för män och kvinnor, 0,39 och 0,23. Dessa värden var inte signifikanta; stickproven var små (11 män och 36 kvinnor), men trenden är den förväntade: högre korrelationer inom könsggrupperna.

I normmaterialet, som omfattade 823 personer i denna analys, ungefär lika många män och kvinnor, var korrelationerna mellan EI och balans signifikanta för båda könsggrupperna, men inte starka: 0,24 ($p < 0,0005$) för män och 0,10 ($p < 0,05$) för kvinnor. När det gäller balans hade männen en tydligt högre nivå än kvinnorna. Fig. 3 visar resultaten efter en gruppindelning av EI-variabeln.

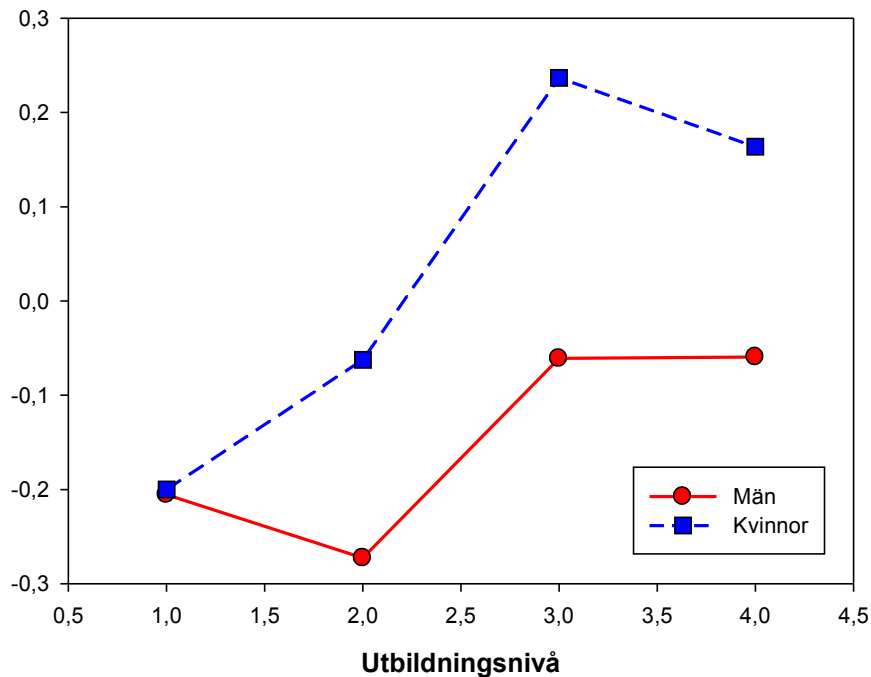
Balans arbete - övrigt liv



Figur 3. Sambandet mellan balans och EI, normmaterialet.

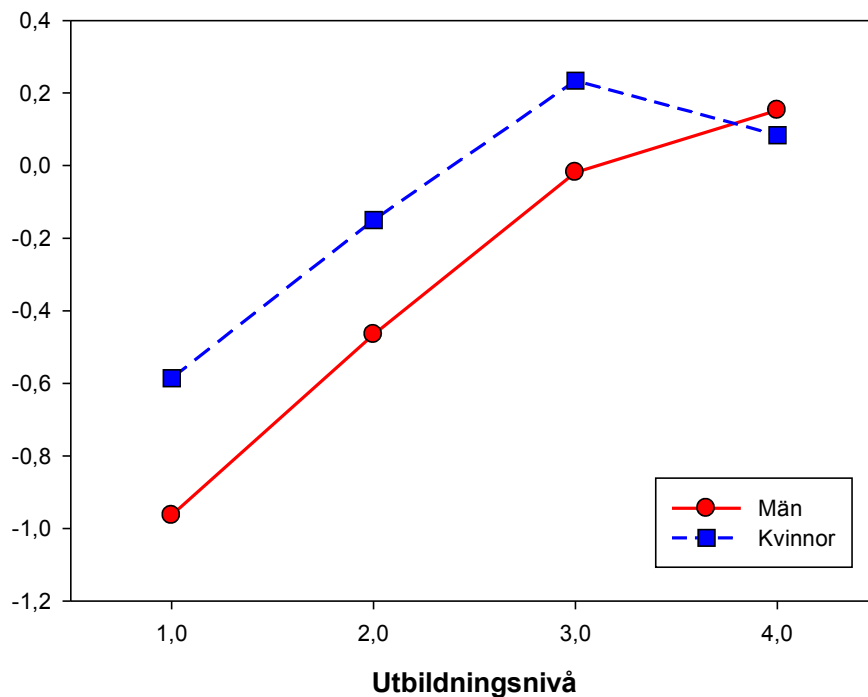
Normmaterialet tillåter också en analys av sambandet med utbildningsnivå, som varierade alltför lite i data från Mälardalens högskola. Det förväntade sambandet var positivt: om testvariablerna tillförlitligt mäter de avsedda begreppen bör både EI och social förmåga öka med utbildningsnivå. Se Fig. 4 och 5.

Emotionell intelligens



Figur 4. Sambandet mellan emotionell intelligens och utbildningsnivå, normmaterialet.

Social förmåga



Figur 5. Sambandet mellan emotionell intelligens och utbildningsnivå, normmaterialet.

Både EI och socialförmåga ökade starkt med utbildningsnivå, både för män och för kvinnor, ett resultat som stödjer validiteten hos dessa mätningar. Emotionell intelligens och social förmåga erhöll alltså stöd i de begreppsvalideringar som gjordes mot ensamhet, kön, utbildningsnivå och civilstånd. Kvinnornas lägre balans var tydlig. Den fanns där trots att

kvinnorna hade högre EI. Troligen beror skillnaden mellan könen i balans på att kvinnorna tar på sig en större del av ansvaret för hem och familj och därigenom utsätts för större konflikter mellan arbete och övrigt liv.

UPP-testet mäter social förmåga. Om denna mätning är tillförlitlig bör testvariabeln samvariera med olika aspekter på framgång.

Social anpassning i arbetslivet mättes med en ny skala, som med framgång kunde begreppsvalideras mot en internationellt etablerad sådan skala. Sambandet med emotionell intelligens enligt testet framgår av Figur 6. Se vidare Rapport 2009:3.

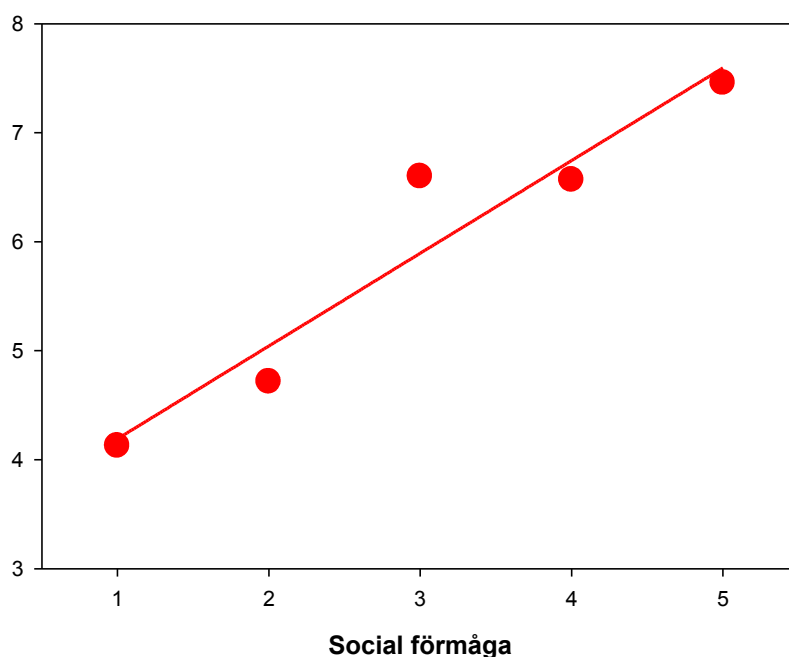
Social anpassning i arbetslivet



Figur 6. Sambandet mellan social anpassning i arbetslivet och emotionell intelligens

Hur är det med social anpassning i livet i stort? De testade fick frågan om hur många riktigt goda vänner de hade, se Figur 7 för sambandet med social förmåga enligt testet.

Antal "riktigt goda" vänner



Figur 7. Sambandet mellan antalet "riktigt goda" vänner och social förmåga.

Dessa resultat visar att testvariabler i *UPP*-testet kan förväntas ha tydliga samband med relevanta aspekter på social och ekonomisk framgång.

Kreativitet

Antalet deltagare i denna delstudie (Studie 8) var 45, 36 män och 9 kvinnor. De arbetade som säljare på ett stort företag inom telekommunikation. Medianåldern var 36 år, variationsvidd 25-56. Tretton hade högskoleutbildning. Alla utom 2 hade minst 2 års erfarenhet av säljarbete, 33 hade mera än 4 års erfarenhet. De hade arbetat på företaget i genomsnitt (median) 3 år, inom branschen i 6 år. Gruppen bestod alltså av personer med ganska omfattande erfarenhet av säljarbete inom branschen. Testning gjordes på uppdrag av företaget som ville använda informationen i sitt HR-arbete. Bakgrunden var en mycket stor variation i arbetsprestation i gruppen.

Två sorters kreativitetstest användes: test på divergent tänkande (4 uppgifter) och självrapportuppgifter (kreativ attityd och öppenhet). Uppgifterna som mätte divergent tänkande var korta (3 minuter var) och de bestod av att skriva ner så många sätt man kunde komma på att använda en 3 decimeter lång linjal eller likheter mellan två till synes mycket olika begrepp som en officersmössa och en roddbåt. De utvärderades efter antalet förslag som skrivits ned utan hänsyn till kvalitet. Kontroller visade att utvärdering med hänsyn till kvalitet gav ett mycket likartat resultat. Självrapportskalorna var ett mått på öppenhet samt skalan som mäter kreativ attityd.

Korrelationer mellan skalorna återges i Tabell 3. Tabellen visar att det fanns höga samband mellan de skilda måtten på kreativitet. Det är särskilt intressant att se att prestationsmåtten på

divergent tänkande samvarierade ganska högt och systematiskt med självrapportskalorna, vilket innebär ett starkt stöd för deras begreppsvaliditet.

Tabell 3. Korrelationer mellan kreativitetsmått.

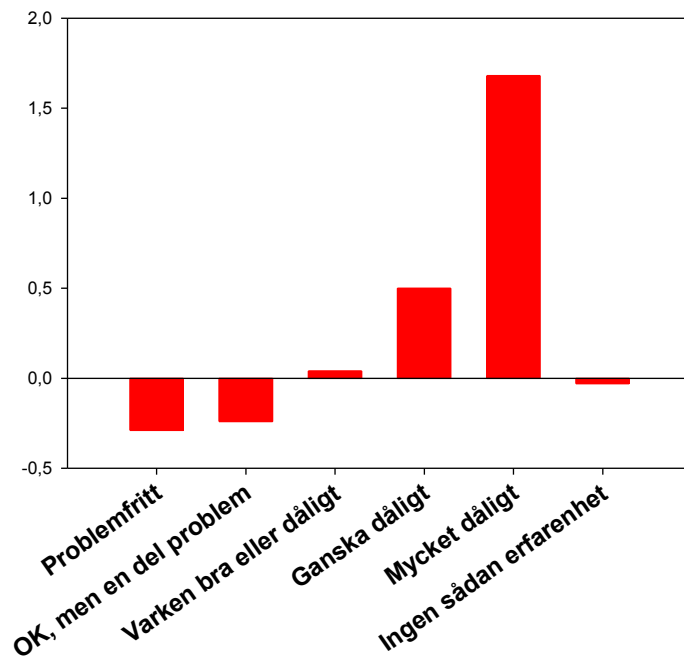
	1.	2.	3.	4.	5.	6.
1. Kreativ attityd	1.00	0.71**	0.49**	0.34*	0.38**	0.36*
2. Öppenhet	0.71**	1.00	0.47**	0.34*	0.38*	0.40**
3. Divergent tänkande 1	0.49**	0.47**	1.00	0.69**	0.49**	0.41**
4. Divergent tänkande 2	0.34*	0.34*	0.69**	1.00	0.70**	0.62**
5. Divergent tänkande 3	0.38**	0.38*	0.49**	0.70**	1.00	0.60**
6. Divergent tänkande 4	0.36*	0.40**	0.41**	0.62**	0.60**	1.00

p<0,05 ** p<0,01

Stresskänslighet

Skilsmässa kan vara en mycket stressande upplevelse. Hur en sådan erfarenhet upplevs borde ha ett starkt samband med testets skala för stresskänslighet. Så var också fallet, i en studie av 107 personer [106], se Fig. 8. Eta-värdet, som motsvarar korrelationen, var 0.49.

Stresskänslighet



Figur 8. Stresskänslighet och upplevelsen av skilsmässa (standardiserad skala)

Bilaga 3. Metodsynpunkter på validering

I denna bilaga diskuteras vilka kriterier som bör användas vid testgranskning. De grundläggande dokumenten är EFPA-kriterierna [59] samt STP:s version av dem³⁵. En diskussion av dessa frågor finns i en epostkorrespondens³⁶ mellan Lennart Sjöberg och David Bartram, ordförande i EFPA:s grupp för testgransningskriterier.

Stickprovets storlek

Stickprovets storlek kan diskuteras. STP ”kräver” $N=100$, annars förklaras valideringen vara baserad på en ”inadekvat” studie., alltså inte inadekvat på något annat sätt. Se STP:s ”Kvalitetskriterier för testinstrument – personlighetsbedömning”, 2002. (Senare upplaga tycks inte finnas). Detta krav tycks vara baserat på EFPA:s kriterier [59], men dessa innebär endast en *rekommendation* om $N=100$ eller större, inte ett *krav*³⁷. Bakom EFPA:s rekommendation ligger i sin tur en ”power-analys”, där man laborerar med följande parametrar:

- Det är önskvärt att kunna belägga ett samband mellan test och kriterium som statistiskt signifikant på nivån 0,05 eller bättre
- Det är önskvärt att kunna belägga sambandet på den nivån om korrelationen i populationen är minst 0,30
- Det är önskvärt att kunna belägga sambandet enligt ovanstående med minst 90 % chans

Analyser av den här typen kan göras med hjälp av tabeller [20]. De tre kraven leder till $N=100$ (som ger chansen att belägga sambandet = 0,86, för att vara mera exakt). Men antag att vi ansätter andra lika rimliga eller lika godtyckliga värden på parametrarna. Om vi vill belägga ett samband utifrån konventionell signifikans med en korrelation på 0,40 i populationen med t ex minst 50 % chans att upptäcka den, vad händer då?

Rekommendationen blir $N=30$ (som ger 61 % chans för att vara exakt). Om man alltså vill pröva hypotesen om en något större korrelation mellan test och kriterium (0,40 i stället för 0,30), och nöjer sig med en mindre men ändå hyfsat stor chans att upptäcka ett samband, sjunker kravet på stickprovets storlek dramatiskt: från 100 till 30.

EFPA:s kriterier är orienterade mot chansen att ”upptäcka” ett signifikant samband. Men vilket evidens värde har ett sådant signifikant samband i betydelsen chansen att det ska kunna replikeras? Killeen³⁸ har föreslagit att man beräknar p_{rep} som är sannolikheten att replikera ett samband eller en differens med samma tecken som i den ursprungliga studien. Om vi har erhållit en korrelation signifikant på nivån 0,05 är sannolikheten att kunna replikera resultatet vid olika storlekar på stickprovet:

³⁵ <http://www.stpsverige.se/>

³⁶ <http://www.psykologisk-metod.se/files/Bartram%202.pdf>

³⁷ Personlig kommunikation, professor David Bartram, 2010-06-09.

³⁸ Killeen, P. R. (2005). An Alternative to Null-Hypothesis Significance Tests. [doi:10.1111/j.0956-7976.2005.01538.x]. *Psychological Science*, 16(5), 345-353.

$N=100, p_{rep} = 0.96$

$N=50, p_{rep} = 0.88$

$N=30, p_{rep} = 0.81$

Chansen är alltså mycket stor att ett resultat visar sig robust även med mycket mindre storlekar på stickprovet än $N=100$. Vid små stickprov finns en risk att inte upptäcka ett samband, men har man väl gjort det är det tämligen pålitligt, med andra ord.

$N=50$ är troligen ett realistiskt krav i många testvalideringar, och som framgår av ovanstående är chansen att få robusta resultat inte mycket mindre än för $N=100$. En studie som visat på ett signifikant samband blir knappast ”inadekvat” av att chansen att sambandet kan replikeras sjunker från 0,96 till 0,88. $N=30$ är en ofta använd minimumnivå i experimentell psykologisk forskning och även den kan försvaras vid testvalidering, ur denna synpunkt.

Det är viktigt att notera att risken att inte få ett signifikant resultat är *forskarens*, inte *användarens*. En misslyckad studie (icke signifikant eller mycket svagt samband) åberopas naturligtvis inte av forskaren/testutvecklaren som stöd för testet. Användaren behöver däremot veta om man kan lita på att ett samband kommer att hålla i nya stickprov. Här kommer p_{rep} in i bilden och ger den information som användaren behöver.

STP har tolkat en *rekommendation* som ett *krav*, med mycket negativa konsekvenser för testbranschen. Det är nämligen i allmänhet mycket svårt, dyrbart och tidsödande att samla in så stora stickprov som $N=100$. Detta torde göra att man hellre avstår. I arbetet med *UPP*-testet har vi samlat in data som i några fall uppfyller kravet $N=100$, i andra inte. Vad dessa data säger om testet beror inte på stickprovets storlek utan på hur starka sambanden är och chansen att de kan replikeras och inte bara är enstaka resultat som därför saknar relevans, och givetvis på om andra krav på studierna uppfylls.

Externa kriterier

För kriterierelaterad validitet krävs externa kriterier. Vad som menas med ”externa” kan emellertid vara oklart. I EFPA:s anvisningar finner man formuleringen ”not part of the instrument”. Kanske ska man därmed förstå att alla testuppgifter som besvaras i samband med testningen avvisas som icke externa. Men det beror i sin tur på vad som menas med ”the instrument” och ”part”.

Proxykriterierna kan troligen anses ej uppfylla kravet på att vara externa i EFPA:s mening, eftersom de onekligen är integrerade i *UPP*-testet. Men antag att de i stället ges i form av ett separat instrument. Ett sådant finns redan. De skulle då inte vara ”part of the instrument”, men resultatet skulle med stor sannolikhet bli detsamma som nu. Validering mot proxykriterier kan därför betraktas som extern. Dessutom tyder våra data på att proxyvalidering ger resultat som är jämförbara med en helt oberoende chefsbedömning av arbetsresultat [110].

I en studie var kriteriet chefsansvar [106]. Detta förefaller vara ett solklart exempel på externt kriterium. I andra lika klara fall handlade det om chefs bedömningar av arbetsresultat. Mera om detta senare.

Korrelationer och andra sambandsmått

EFPA/STP bygger sina granskningskriterier till stor del på ”korrelationer” mellan testvariabler och kriterier. Detta gäller både begreppsvalidering och validering mot externt kriterium. Med korrelationer verkar man mena vanliga Pearsonkorrelationer som ju förutsätter att man har två variabler som mäts i flera steg och som är relaterade med linjära regressioner. EFPA-kriterierna specificerar vilka korrelationsnivåer som ska uppnås.

Det finns emellertid många sammanhang när detta synsätt är otillräckligt. En eller flera av variablerna kan vara binär (t ex chef/ej chef) och då bör vanliga korrelationer inte beräknas. Korrelationen mellan en binär variabel och en flerstegsvariabel brukar kallas punktbiserial och den ger ett lägre värde än vad som erhålls om den binära skalan är en flerstegsvariabel också den. Den biseriala korrelationen är en enkel lösning på detta problem, eller binär logistisk regression, vilket har använts i en av studierna i denna manual [106]. I det senare fallet finns metoder för att uppskatta korrelationen, men det är alltså inte längre frågan om en konventionell och rättfram korrelationsberäkning. Om den beroende variabeln är en klassifikation i flera steg än två krävs andra typer av analyser. Om den består av ordnade kategorier kan man använda polyseriala korrelationer under antagande om underliggande normalfördelning. Dessa blir normalt högre än vanliga korrelationer på rådata som antagit intervallskalor. De används ofta vid SEM-analyser (LISREL etc).

I vissa fall kan sambanden vara kurvilinear. Då kan vanliga korrelationer inte användas eftersom de underskattar sambandets styrka. Man måste anpassa andra regressionsfunktioner än linjära sådana. Även i dessa fall kan man uppskatta andelen förklarad varians i den beroende variabeln.

Det finns alltså många fall i praktiskt arbete som kräver andra typer av analyser än enkla korrelationer, och där korrelationer kan uppskattas under vissa antaganden, eller där förklarad variansandel kan uppskattas. Om man använder EFPA/STP-kriterierna finns en risk att sådana analyser ignoreras eftersom de inte på ett enkelt sätt låter sig inordnas i konventionellt korrelationstänkande. David Bartram skriver: “As to your questions about construct validity then I would say that any relevant measure can be used – we should not be restricted to correlations. In general one is looking at the magnitude of measures of association – whether Pearson r, eta or anything else – or indeed any other relevant indicator of effect size (e.g. d values).” (Personlig kommunikation 2010-06-10).

Mätfelens inverkan på korrelationer

Bakom de observerade, manifesta, värdena i alla mätvariabler antar vi att det finns latent dimensioner som är av intresse. De manifesta variablerna är emellertid också påverkade av mätfel. I begreppsvalidering vill vi uppskatta sambandet mellan latent variabler, inte manifesta.

Mätfelen gör att sambanden mellan variablerna blir svagare än vad sambanden är mellan motsvarande latent variabler. Om man har värden på variablernas reliabiliteter, och detta är ofta fallet, kan man korrigera för mätfel och beräkna sambanden mellan de latent variablerna. Detta är den relevanta analysen vid begreppsvalidering, men den nämns inte av STP. Om man ignorerar mätfelens betydelse och enbart intresserar sig för samband mellan manifesta variabler måste detta motiveras, men det gör inte STP. David Bartram skriver: “When considering ratings, reviewers should take into account whether they are looking at corrected or uncorrected correlations and bear in mind that uncorrected value underestimate true validity”. (Personlig kommunikation, 2010-06-10).

Summering av ett tests validitet i fallet med flera underordnade testskalor eller flera kriterier

EFPA talar endast om medianer av korrelationer mellan testet och kriterier. Men hela testet kan ha en hög validitet utan att mer än några få av testvariablerna har det. Det är i själva verket troligt att så ofta blir fallet eftersom olika testvariabler är valida för olika kriterier, t ex av socialt eller uppgiftsorienterat slag [43]. EFPA-dokumentet, som anger hur utvärderingen av testet ska göras, tar inte explicit upp denna frågeställning utan talar om "testets" validitet. Om ett test mäter flera dimensioner än en, vilket torde vara det absolut vanligaste, måste man rimligen med "testets" validitet avse validiteten hos hela testet simultant – t ex i en multipel regression - inte ett genomsnitt av validiteterna för de enskilda testvariablerna. Jämför med stora meta-analyser som [53]. Om stickprovet är litet är en multipel regression olämplig och man kan t ex vikta de oberoende variablerna lika. En liknande summering av kriterier kan också vara lämplig, för att ge en överskådlig bild av validiteten.

Slutsatser

Ett antal viktiga punkter av betydelse för testgranskning sammanfattas här.

1. Inget krav på stickprovsstorlek vid validering har rationell grund, och ställs inte heller av EFPA. Signifikanta resultat är relevanta även om $N < 100$.
2. Begreppsvaliditet bör utvärderas efter korrektion för mätfel.
3. Korrelationer är bara ett av många sambandsmått som är relevanta och användbara vid testgranskning. Linjära korrelationer är olämpliga i vissa sammanhang.
4. Ett tests validitet bör utvärderas för hela testet simultant, inte som en median av korrelationer mellan enskilda testskalor och ett eller flera kriterier.
5. Begreppsvalidering är den viktigaste typen av validering.
6. Med externa kriterier bör man avse all relevant kriterieinformation som inte är en inherent del av testskalorna själva. Informationskällan kan vara den testade själv som uppger t ex sin utbildningsnivå, "objektiva" data eller bedömningar av chefer.

Bilaga 4. Kompletterande information om reliabilitet

Reliabilitetsvärden för skalorna i UPP-testet.			
Skala	Antal uppgifter eller ingående variabler för de aggregerade måtten	Reliabilitet (homogenitet)	Genomsnittlig korrelation mellan uppgifterna som ingår i skalan
Utåtvändhet	10	0.79	0.28
Vänlighet	8	0.78	0.31
Emotionell stabilitet	8	0.71	0.23
Öppenhet	9	0.63	0.16
Noggrannhet	8	0.74	0.24
Uthållighet	8	0.79	0.31
Samarbetsvilja	10	0.75	0.23
Positiv grundhållning	10	0.85	0.36
Självförtroende	9	0.75	0.25
Social förmåga	8	0.65	0.19
Emotionell intelligens, självrapport	18	0.84	0.22
Emotionell intelligens, emotionsbedömningar	4 (32)	0.52	0.23
Kreativitet	8	0.73	0.31
Perfektionism	8	0.69	0.19
Arbetsstillfredsställelse	3	0.92	0.79
Arbetsvilja	7	0.88	0.52
Resultatorientering	13	0.73	0.17
Förändringsvilja	8	0.54	0.13
Arbetsintresse	6	0.71	0.29
Balans arbete – övrigt liv	9	0.85	0.39
Skön målning 1 (kovert)	12	0.76	0.21
Skön målning 2 (overt)	9	0.70	0.21
Hedonisk ton	4	0.76	0.44
Aktivitet	4	0.84	0.56
Spänning	4	0.78	0.47
Humör, sammanslaget	12	0.85	0.32
Attityd till testet	8	0.79	0.32
Jagstyrka, korrigerade data	11 variabler	0.75	0.21
Chefspotential, korrigerade data	13 variabler	0.76	0.20
Stresskänslighet, korrigerande data	6 variabler	0.60	0.20
Median		0.76	0.25

Källa: [110]

Psykometriska egenskaper hos om skalorna i screeningmodulen. N=3486.				
Skala	Reliabilitet (Cronbachs alfa)	Antal uppgifter	Genomsnittlig korrelation mellan uppgifterna i skalan	Andel av variansen som förklaras av skönmålning
Uthållighet	0,65	8	0,19	0,183
Samarbetsvilja	0,83	11	0,30	0,409
Kreativitet	0,68	8	0,20	0,060
Utåtvändhet	0,81	10	0,29	0,326
Positiv grundattityd	0,76	10	0,25	0,354
Kovert skönmålning	0,69	12	0,16	-

Källa: [111]