

SSE/EFI Working Paper Series in Business Administration  
No 2012:2

# Skönmålning i personlighetstest

**Lennart Sjöberg**

**Reviderad Juni 2013**

Stockholm School of Economics  
SSE Institute for Research  
Center for Risk Research  
and  
Center for Media and Economics Psychology  
Marketing and Strategy Department

## Innehåll

Förord .....	3
Abstract .....	4
Executive summary .....	5
Inledning.....	9
Terminologi.....	9
Skönmålning som aktuellt forskningstema .....	10
Historik.....	10
Mätning av skönmålning.....	11
Social önskvärdhet och självbedrägeri .....	11
Validering av korrektion för skönmålning .....	14
Personlighet eller skönmålning? .....	19
Svarstid.....	20
Hur vanlig och stark är skönmålning?.....	22
Vad är det som ger upphov till skönmålning?.....	25
Skönmålning och teststruktur.....	25
Metodik för att handskas med skönmålning .....	25
Suppressormodellen för korrektion för skönmålning.....	25
Normativ metodik .....	25
Ipsativa svarsformat .....	29
Användning av ett gränsvärde.....	31
Varningar.....	32
Lura testet? .....	32
Ordningen mellan testuppgifterna.....	34
Effekter av skönmålning .....	35
Skönmålning och femfaktormodellen .....	36
Normering .....	36
Upprepade testningar och skönmålning .....	36
Debatten om skönmålning.....	37
Diskussion och slutsatser .....	38
Uppsummering .....	40
Referenser.....	41
Appendix 1: Marlowe-Crownes skala och två skalor som mäter "Need for approval" .....	47
Appendix 2. Korrektion för skönmålning: metodik .....	50

## Förord

Syftet med denna rapport är att sammanfatta forskningen och debatten om skönmålning ("faking") i personlighetstest. Skönmålning är ett stort problem, både enligt användare av test och forskare på området, även om dessa grupper troligen tenderar att se olika på problematiken. Praktiker inom HR-området, liksom ansvariga chefer i stort inom näringslivet, är ibland skeptiska till värdet av personlighetstest därför att de anser att resultaten på sådana test kan snedvridas av att de testade medvetet ger en förskönad bild av sig själva. Mot detta står vissa forskare och testleverantörer som ibland hävdar att skönmålning knappast förekommer, eller – vanligare - att skönmålning inte har allvarliga effekter på testens validitet, eller att det finns effektiva metoder för att förebygga skönmålning eller korrigera för den. Vem har rätt?

Det är helt klart att skeptikerna har rätt, när det gäller alla de test som är helt eller nästan helt oskyddade mot effekterna av skönmålning. I denna rapport ska jag visa att det finns en effektiv metodik för att korrigera för dessa effekter.

Stockholm i november 2012

I föreliggande version har diskussionen om forskningen på området uppdaterats och utökats. Vissa nya analyser av egna data har också lagts till.

Nerja i maj 2013

## Abstract

Skönmålning är ett stort problem vid användning av personlighetstest av självrapporttyp. Det är ofta lätt att genomsåda vad olika testfrågor syftar till att mäta och vill man bluffa är det lätt att göra det. Trots det är denna typ av test vanliga och de tycks få en ökande användning i näringslivet. Det kan delvis bero på att vissa test använder ett format (ipsativt) där den testade är instruerad att välja mellan alternativ som man har försökt matcha i social önskvärdhet. Sådana test har antagligen en viss trovärdighet, men forskningen visar att de inte lyckas särskilt väl med att eliminera effekterna av skönmålning. Dessutom samvarierar de med intelligens, och den validitet de eventuellt kan ha beror troligen till viss del på detta. Ett annat skäl till den omfattande användningen av personlighetstest är det vanliga påståendet att dessa test, även om de är helt oskyddade mot skönmålning, har högt prognosvärde i arbetslivet. Men detta påstående är felaktigt om man ser till den omfattande forskningen om "Big Five" som visat att dessa dimensioner bara har en marginell prognosförmåga i förhållande till arbetsresultat. Ett test som är oskyddat mot skönmålning, eller bara mycket ineffektivt skyddat, ger regelmässigt kraftiga överskattningar av testvärden hos dem som utnyttjar chansen att svara taktiskt. Andra, som kvinnor och invandrare, skönmålar i genomsnitt mindre och missgynnas därför av testen. I denna rapport beskrivs en alternativ metodik som bygger på användning av särskilda skalor för att mäta tendensen till skönmålning. Modeller som anpassas på grundval av empiriska data och som är olika för olika testskalor eliminerar ca 95 % av effekterna av skönmålning. Metodiken har validerats i en rad empiriska studier som kortfattat beskrivs i rapporten. Korrektion för skönmålning är av mycket stor betydelse vid tolkning av testdata, och för etablering av en rättvisande normdatabas.

## Executive summary

Personlighetstest har blivit vanliga både inom privata näringslivet och inom offentlig sektor. Det gäller både vid urval för anställning och vid utvecklingsarbete. Det är en positiv trend, men den är inte utan problem. Många test marknadsförs med hjälp av okritiska lovprisanden, samt bilder på glada och vackra människor på hemsidan, och man håller tyst om den eventuella forskning som gjorts. Manualer och rapporter behandlas som ”affärshemligheter” och är inte tillgängliga för den som vill bilda sig en uppfattning om ett test utan att först köpa utbildning och licens (som kan sammanlagt kosta upp till SEK 40 000). Kundernas intelligens och intresse av kritisk utvärdering underskattas. Skulle man köpa en bil om man fick teknisk information om den först när man köpt den?

Personlighetstesten är oftast av självrapporttyp. Det handlar alltså om ett antal påståenden eller frågor (item) där den testade uppmanas att ta ställning till vad som bäst beskriver honom eller henne. Det är ofta lätt att genomsåda vad olika testfrågor syftar till att mäta och vill man bluffa är det lätt att göra det, att skönmåla. Skönmålning innebär att man ger en överdrivet positiv bild av sig själv och det är ett av de stora problemen vid användning av personlighetstest. Självrapporttest är vanliga, trots problemet med skönmålning. Aktuella uppskattningar av förekomsten av skönmålning vid urval till anställning visar emellertid att minst 30 % av de testade skönmålar, troligen fler. Behovet av metodik för att motverka skönmålning eller – helst – att eliminera dess effekter är uppenbart. Det visar sig också att många testleverantörer använder någon form av skalor för att mäta skönmålning, även om de gör det mindre effektivt än vad som är möjligt. Det vore märkligt om de inte gjorde detta på goda grunder och den erfarenhet de har, oskyddade test ”slår i taket” – många testade får extremt höga (och orealistiska) resultat som inte är trovärdiga eller särskilt användbara.

Vissa testleverantörer har försökt ”lösa” detta problem genom att förneka att det existerar. Olika argument har framförts:

- ”Skönmålning förekommer knappt”. Helt fel, skönmålning är vanlig och ofta stark.
- ”Det finns inget 100 % säkert sätt att upptäcka vilka som har skönmålat”. Naturligtvis inte, i psykologin är ingenting 100 % säkert! Poängen är att man kan komma en bra bit på väg med de metoder som används.
- ”Testens validitet (”träffsäkerhet”) ökar inte om man korrigerar för skönmålning, den kan t o m minska”. Detta är sant ibland, men ibland ökar den. Effekterna är dock små; detta sätt att utvärdera metodik för korrigerande är okänsligt. Det viktiga är att den lilla gruppen av de testade som till slut kommer ifråga för anställning tenderar att innehålla många bluffare, om man använt oskyddade test. Vem vill anställa bluffare till viktiga arbetsuppgifter?
- ”Den som skönmålar vet vad som krävs i jobbet och får han eller hon det, så är det bara att fortsätta att skönmåla, och alltså att göra ett bra jobb”. Detta något bisarra argument förväxlar förmåga att göra något med kunskap om vad som krävs för att lyckas. Helt olika saker i många sammanhang.
- Skalor för mätning av skönmålning mäter egentligen ”personlighet” – man kan och bör inte avlägsna den komponenten från testet. Men att skönmålning korrelerar med testskalorna är ett uttryck för att dessa är mer eller mindre laddade med just skönmålning, inte tvärtom. Hur skulle fö en skala för skönmålning kunna mäta ett stort antal personlighetsaspekter, och varför bara somliga och inte andra?

- Ett ofta diskuterat specialfall av det föregående argumentet är att traditionella skalor för mätning av ”social önskvärdhet” i själva verket mäter personlighetsvariabeln ”behov av att vara omtyckt”. Det visar sig emellertid att mera direkta mått på denna variabel inte har något samband med skalor av den traditionella typen.
- Korrektion för skönmålning innebär att man bildar differenser, och sådana ska enligt testteori vara mindre reliabla (mera behäftade med mätfel) än de ingående grundläggande skalorna. Därför är det bara att vänta att gruppskillnader försvinner när man har korrigerat – om mätfelen är stora är det ju bara slumpfel kvar och dessa ska enligt teorin inte ha samband med några andra variabler. Detta argument utgår från att korrektionen medför mycket stor sänkning av reliabiliteten, men analys har visat att den visserligen förekommer men är av obetydlig storlek.
- Man kan helt enkelt varna de testade för att skönmåla; det borde eliminera problemet. Forskning om denna ansats visar emellertid att den har små och inkonsistenta effekter.
- Man använder testet för ”screen out”, dvs gallra bort den del av de sökande som har lägst värde. Bland dem finns det troligen få som har skönmålat. Men problemet med de övriga kvarstår. Detta är ett ineffektivt sätt att använda testresultatet.

Ett ganska vanligt sätt att försöka hantera skönmålning är användning av ett format (ipsativt eller jämförande format) där den testade är instruerad att välja mellan alternativ som man har försökt matcha i social önskvärdhet.

Ett exempel kan vara: Vilket passar bäst in på dig av följande två påståenden:

- A. När jag ska tala inför en grupp känner jag mig lugn och samlad.
- B. Jag får lätt många nya idéer.

Test som använder sådana format har antagligen en viss trovärdighet, men forskningen visar att de inte lyckas särskilt väl med att eliminera effekterna av skönmålning. En stor del av effekterna finns kvar. Dessutom samvarierar de med intelligens, och den validitet de eventuellt kan ha beror troligen till viss del på detta.

Det finns ännu flera problem med ipsativa test. De testade tycker inte om ipsativa format. De tycker att det är svårt att ta ställning, att ”jämföra äpplen och päron”. Det tar också längre tid att besvara sådan item. De har inte högre validitet än andra test, tvärtom kan de leda mycket fel eftersom de *i princip* inte tillåter jämförelser mellan individer. Alla data uttrycker ju bara jämförelser inom varje testad person. Man kan t ex vara mera utåtvänd än kreativ, men trots det inte särskilt utåtvänd. För en person med höga värden i många av de personlighetsegenskaper som mäts i ett ipsativt test, kan låga resultat i några skalor helt enkelt avspegla att han eller hon ligger något lägre i dessa, men ändå på en hög nivå. En 10-kampare som ligger 1:a i 9 grenar men 2:a i höjdhopp är bra även i denna gren, men ipsativt sett är han en dålig höjdhoppare.

En annan ganska vanlig metodik är att mäta skönmålning i en grupp av testade personer, sedan gallra ut dem som har höga värden, t ex de 25 % som har de högsta värdena. Man finner emellertid efter en noggrann analys att denna metodik är mycket ineffektiv. En mycket stor del av effekterna av skönmålning finns kvar även efter gallringen.

Ett skäl till den omfattande användningen av personlighetstest är det vanliga påståendet att dessa test, även om de är helt oskyddade mot skönmålning, har högt prognosvärde i arbetslivet. Men detta påstående är felaktigt om man ser till den omfattande forskningen om

"Big Five", ledande i många tillämpningar, som visat att dessa dimensioner bara har en marginell prognosförmåga i förhållande till arbetsresultat. I termer av korrelation mellan testskala och arbetsresultat handlar det om högst 0.3, ett blygsamt resultat som inte är bättre än vad man uppnådde redan på 1920-talet. Siffran 0.3 gäller dimensionen noggrannhet/pålitlighet, medan de övriga 4 dimensionerna ligger mycket lägre, i flera fall på ca 0. Det kan finnas många skäl till dessa svaga resultat, ett är troligen att "Big Five"-dimensionerna är alltför allmänna och saknar fokus på arbetsresultat, ett annat att man inte kan hantera skönmålningen. I skarpa lägen, inför viktiga beslut om t ex anställning, överdriver många sina "meriter". Detta gäller i synnerhet män, varav många passar på att framställa sig själva som ytterst framstående. Kvinnor ger mera realistiska bilder av sig själva och missgynnas därför av test som inte hanterar skönmålning. Samma gäller invandrare jämfört med testade personer av svensk härkomst.

I denna rapport beskrivs en metodik för att hantera skönmålning som bygger på användning av särskilda skalor för att mäta tendensen till skönmålning. En sådan skala, som utvecklats för UPP-testet, bygger på en traditionell ansats för att mäta "social önskvärdhet. Den innehåller påståenden om beteenden som är vanliga men ej önskvärda ("Jag har aldrig tyckt intensivt illa om någon") eller önskvärda men ovanliga ("innan jag röstar läser jag noga alla partiernas program"). Denna skala har vi kallat "overt" och den kan eventuellt genomskådas. Därför finns i UPP även en "kovert" skala som är mycket svår att genomskåda och som innehåller den vanliga typen av påståenden (item) i ett personlighetstest, men item som är utvalda för att de har speciellt höga samband med den overta skalan.

Modeller som anpassas på grundval av empiriska data och som är olika för olika testskalor eliminerar ca 95 % av effekterna av skönmålning. Detta har jag funnit i flera studier, bl a en av sökande till officersutbildning vid Försvarshögskolan. Där hade vi möjlighet att studera ett flertal grupper som tog testet under varierande grad av "skapt läge", från dem som redan var antagna (minst skarphet) till dem som uppfattade att testresultaten skulle påverka deras chans att bli antagna. Graden av skönmålning varierade som väntat kraftigt som en funktion av hur skarpt läget var vid testningen. Resultaten i personlighetsskalor varierade likaså, trots att grupperna var likvärdiga. Korrektion för skönmålning enligt vår metodik eliminerade skillnaderna mellan grupperna till ca 95 %.

Sådana data stöder metodiken och antagandena bakom den. Även andra resultat ger stöd för metodiken. Tiderna (i sekunder) för att besvara varje item registreras i UPP-testet. Vi har funnit att personer som skönmålar mera tar mera tid på sig för att besvara testuppgifterna, och detta var väntat. De måste fundera på hur de ska svara för att vara taktiska. Vi har också helt enkelt frågat testpersonerna om de skönmålade. Det fanns ett tydligt samband mellan svar på den frågan och graden av skönmålning utom för gruppen som absolut förnekade att de skönmålade. De hade en hög grad av skönmålning, vilket är rimligt med tanke på att deras förnekande också var en form av skönmålning.

Korrektion för skönmålning är av mycket stor betydelse vid tolkning av testdata, och för etablering av en rättvisande normdatabas. När det gäller tolkning kan man hamna helt fel om man tar okorrigerade testresultat för vad de tycks säga på ytan; många ger en överdrivet positiv bild. Något liknande kan sägas om normer, som är den för ett test livsviktiga jämförelsegruppen. Om data för normgruppen insamlats i oskarpt läge, vilket ofta är fallet, och om inget gjorts för att mäta skönmålning blir normen felaktig och de som testats i en praktisk tillämpning får av detta skäl alltför höga värden. Det finns ett antal exempel på irrelevanta normer från testföretag som ignorerar hela frågan.

Slutsatser:

1. Förekommer skönmålning, är den vanlig och stark? Svaret är ett entydigt ja.
2. Har skönmålning effekter på beslutsfattandet på grundval av testning, t ex vid urval? Svaret är ja.
3. Här skönmålning betydelse för normdata? Svaret är ja, om testet ska användas i skarpt läge måste normdata vara insamlade i skarpt läge.
4. Kan skönmålning undvikas, åtminstone till betydande del, med hjälp av
  - a. Varningar
  - b. Omformulering av item till att bli mera neutrala
  - c. Särskilda mätningar av tendensen att tillskriva sig mera kunskaper än man har
  - d. Anpassning av matematiska modeller till testdata och mätning av i vilken mån testpersoner avviker från dessa modellerSvaret för alla dessa ansatser är nej, de löser inte problemet
5. Ett ganska vanligt sätt att hantera skönmålning tycks vara att utesluta de testpersoner som har högst värde i en sådan skala, t ex de 25 % som ligger högst. Detta har en svag och otillräcklig effekt, se Reeder och Ryan (Reeder & Ryan, 2012).

Ännu tycks det inte finnas någon ny metodik för att upptäcka och korrigera för skönmålning som är tillräckligt lovande för att leva upp till kraven att den ska vara teoretiskt och empiriskt väl underbyggd, och praktiskt användbar. Korrektur med en särskild skala som mäter skönmålning fungerar, även om inte heller denna metod helt eliminerar effekterna.



## Inledning

Personlighetstest av självrapporttyp har funnits sedan 1920-talet (Gibby & Zickar, 2008) och de dominerar testmarknaden i svenskt arbetsliv. Användningen tycks vara ökande. Sådana test är även mycket vanliga i andra länder som USA och Storbritannien och därifrån kommer det mesta av den forskning som jag kommer att diskutera i denna rapport.

Trots expansionen av personlighetstestningen finns det en betydande skepsis mot sådana test, kanske främst beroende på den utbredda uppfattningen att man kan bluffa på dem (Rees & Metcalfe, 2003). Det är en uppfattning som delas av många potentiella testanvändare och testpersoner. Testen förväntas alltså inte ge en rättvisande resultat, vilket givetvis är allvarligt för dem som avser att använda dem som en del av underlaget för beslut. Det är också allvarligt för dem som testas eftersom de upplever det som orättvist om vissa kan bluffa sig fram till en god placering i en grupp av sökande.

Alternativ som projektiva test har sina egna svagheter (Lilienfeld, Wood, & Garb, 2000), som jag inte ska diskutera här. Ibland kan ett intressant alternativ vara rapporter från personer som väl känner den vars personlighet vi är intresserade av (Vazire, 2006), dock sällan i praktiska tillämpningar av personlighetsmätning vid urval. Inom attitydmätning har man utvecklat en intressant metodik som bygger på reaktionstider och som tycks kunna mäta attityder som anses vara fördomar som människor ogärna vill medge att de har (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). Man tycks dock kunna skönmåla även i sådana test, eller åtminstone kunna lära sig det (Röhner, Schröder-Abé, & Schütz, 2011).

I denna rapport ska jag visa att det finns validerad metodik för att komma förbi skönmålningen i självrapporttest, trots att den i viktiga sammanhang är både vanlig och stark.

## Terminologi

Några terminologiska frågor ska först utredas.

**Skönmålning.** Med denna term avses en tendens hos en testad person att ge en överdrivet positiv bild av sig själv i testet. I undantagsfall förekommer motsatsen, alltså att den testade ger en alltför negativ bild av sig själv (Boon, Gozna, & Hall, 2008; Eitan, 2011), men detta fall diskuteras inte i denna rapport. I den internationella litteraturen används ofta ordet "faking". Men även andra termer som "impression management" (skönmålning) och "self-enhancement" förekommer.

Jag har föredragit att använda en neutral terminologi så långt det gått. Men området är fullt av starka åsikter som driver en del skribenter till retoriska ordval. I retoriskt syfte används ord som fusk ("kan vi verkligen anklaga människor för att ha fuskat?") eller lögn ("vi kan väl inte vara säkra på att människor ljugar").

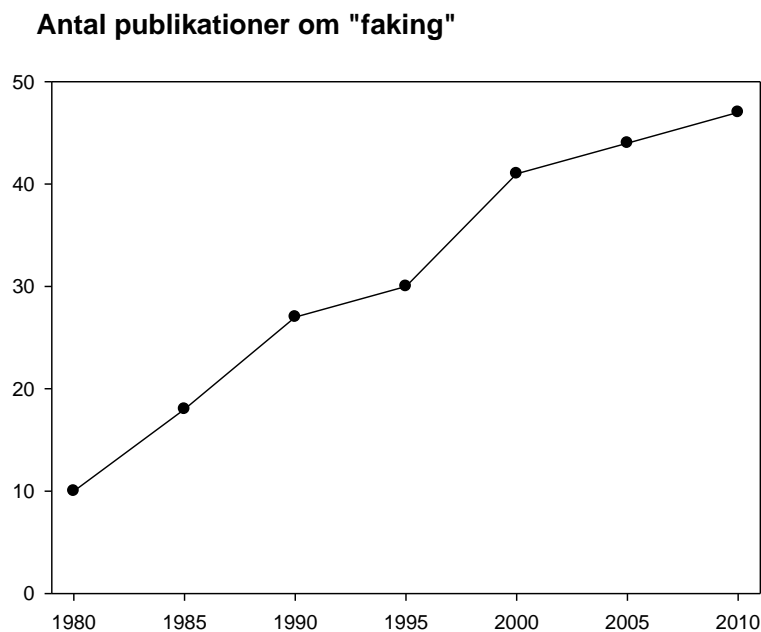
**Självrapport.** Härmed avses att testet är uppbyggt av frågor/påståenden som ska besvaras utifrån den uppfattning den testade har om sig själv och sitt beteende, vanor och erfarenheter etc. Svaren på sådana frågor sammanställs till personlighetsskalor. Naturligtvis är dessa

självrappporter inte alltid sanna; hela problematiken med skönmålning är ju just denna. Om man så vill kan man tala om självpresentationer i stället för självrapporter – det gör faktiskt ingen skillnad. I testet skapar och kommunicerar den testade en bild av sig själv; den kan vara sann eller helt eller delvis falsk.

## Skönmålning som aktuellt forskningstema

Det finns en mycket livaktig forskning om skönmålning. Den dokumenteras vid internationella konferenser, i böcker och tidskrifter som *Journal of Applied Psychology*, *Personnel Psychology* och *International Journal of Selection and Assessment*. En utmärkt källa är en bok, som främst innehåller översikter av forskningen om olika aspekter på skönmålning och som har titeln *A closer application of applicant faking behavior* (Griffith & Peterson, 2006). En senare bok redigerad av Ziegler, Maccann och Roberts (Ziegler, Maccann, & Roberts, 2012) ger en bra komplettering.

Fig. 1 visar antalet publicerade artiklar i databasen PsychINFO funna med sökordet ”faking”. Figuren visar ett starkt ökat intresse bland forskarna sedan 1980. En sökning på Google 2012-09-06 med sökorden ”faking personality tests” gav ca 2 760 000 träffar.



**Figur 1. Antal vetenskapliga artiklar om "faking" publicerade 1980 - 2011**

Figuren visar på ett starkt och ökande intresse bland forskarna för skönmålning. Detta tyder på att problematiken är högst verklig, och ingalunda icke-existerande eller löst för länge sedan.

## Historik

En tidig studie av Hendrickson visade att testpersoner kunde förbättra sina resultat på ett personlighetstest (Hendrickson, 1934). Många senare forskare har funnit liknande resultat (Zickar & Gibby, 2006). En vanlig uppskattning är att 30-50 % av de sökande till ett jobb skönmålar (Donovan, Dwight, & Hurtz, 2003; Griffith & Converse, 2012a). Denna forskning

har oftast byggt på jämförelser mellan testningar under olika instruktioner, som att "svara ärligt" och "försök få ett så bra resultat som möjligt" (Griffith, Malm, English, Yoshita, & Gujar, 2006). Ibland har det varit samma personer som tagit testet under dessa två typer av instruktioner vid två tillfällen. Skillnaden mellan tillfällena tas som ett mått på skönmålning (Ellingson, Sackett, & Hough, 1999). Ett problem med denna metodik är att skönmålning i skarpt läge tycks vara något väsentligt annorlunda än skönmålning under instruktionen att framstå som "bra" i en experimentell upplägning (Pauls & Crost, 2005). En annan metodik är att mäta skönmålning vid varje enskilt testtillfälle, men det verkar som om måtten på skönmålning inte är jämförbara (Ferrando & Anguiano-Carrasco, 2011). Skillnaden mellan "ärliga" svar och svar som getts för att ge en positiv bild tyder på en mycket större effekt av skönmålning än mätning av skönmålning vid ett testtillfälle för varje person (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Viswesvaran & Ones, 1999). Detta innebär att enstaka mätning ger en underskattning av graden av skönmålning.

En traditionell metod för att minska eller eliminera effekterna av skönmålning har varit att använda s.k. ipsativa svarsformat. Sådana svarsformat kräver av den testade personen att han eller hon jämför två svarsalternativ som är matchade i social önskvärdhet och väljer det som passar bäst som beskrivning av den egna personen. En annan metod är att göra separata mätningar av tendensen att svara "socialt önskvärt" och sedan korrigera svaren på testet utifrån dessa mätningar. Dessa två och även visa andra sätt att handskas med problemet kommer att behandlas i denna rapport.

## **Mätning av skönmålning**

### ***Social önskvärdhet och självbedrägeri***

Den enklaste metoden för att mäta skönmålning går tillbaka på Marlowe-Crownes skala för social önskvärdhet (Crowne & Marlowe, 1960). Sedan den publicerades för ca 50 år sedan har den använts i 1000-tals undersökningar. Den bygger på idén att en testperson som inte vill medge att han utfört vissa vanliga, men socialt negativa, handlingar troligen skönmålar. Detsamma gäller om en person som påstår att han utfört ovanliga med socialt önskvärda handlingar.

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. Jag kommer senare i rapporten att redogöra för resultat som stödjer det påståendet.

Paulhus har fört vidare diskussionen om skönmålning på ett intressant sätt (Paulhus, 1991; Paulhus & Reid, 1991). Han använde samma typ av uppgifter som Crowne och Marlowe 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 troligen för sig själv. Ett exempel är:

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

Lönnqvist et al. (Lönnqvist, Paunonen, Tuulio-Henriksson, Lönnqvist, & Verkasalo, 2007) fann att Paulhus-skalorna åtminstone till en del tycktes mäta skönmålning 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. Hur som helst korrelerar de ganska högt med den traditionella typen av taktiksvår. I en urvalssituation (N=282, sökande till Handelshögskolan i Stockholm) användes Marlowe-Crownes skala och Paulhus båda skalor. Hans skala som byggde på frågor om normbrott korrelerade 0.84 med Marlowe-Crownes skala, medan skalan på "självbedrägeri" korrelerade 0.54 med Marlowe-Crownes skala – också det ett högt värde.

Är skönmålning medveten eller omedveten? En ganska vanlig uppfattning tycks vara att skönmålning av traditionell typ (Marlowe-Crowne) är ett medvetet försök till skönmålning medan självbedrägeri skulle vara omedvetet. Det tycks dock saknas forskning som ger ett direkt stöd till denna tolkning.

Ett annat viktigt problem är frågan om uppgifterna i en skönmålningsskala går att genomsöka. 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 testuppgifter 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.

I arbetet med *UPP*-testet har konstruerats en egen skala enligt samma principer som Marlowe-Crownes och Paulhus skalor, en overt skönmålningsskala (öppen mätning), men också en kovert skala (dold mätning)<sup>1</sup>. Den senare består av uppgifter av samma typ som i den övriga delen av testet, men de är särskilt utvalda för att korrelerade högt med den overta skalan för skönmålning. På så sätt får man en möjlighet att mäta taktiksvår som inte kan avslöjas av sofistikerade testtagare, något som annars kan vara ett problem (Hurtz & Alliger, 2002).

De två skönmålningsskalorna (overt och kovert) har begreppsvaliderats, se Tabell 1. Båda korrelerade högt med Marlowe-Crownes skala och högt sinsemellan. Den senare korrelationen var något lägre, vilket är rimligt.

**Tabell 1. Korrelationer mellan de tre skönmålningsskalorna (N=159), data från Studie 1 i UPP-testets tekniska manual (Sjöberg, 2011).**

	Marlowe-Crownes skala	Overt skönmålnings-skala	Kovert skönmålnings-skala
Marlowe-Crownes skala	1.00	0.76	0.73
Overt skönmålningskala	0.76	1.00	0.56
Kovert skönmålningskala	0.73	0.56	1.00

För att jämföra effekterna av de två skalorna vid korrektion för skönmålning gjordes sådana korrektioner separat för kovert och overt skönmålning, sedan med båda samtidigt. Data kom från testningar i skarpt läge av 1090 chefskandidater inom både offentlig sektor och det

<sup>1</sup> En intressant fråga för fortsatt forskning är om den koverta skalan mäter omedveten skönmålning, och/eller självbedrägeri.

privata näringslivet. Korrelationen mellan data korrigerade med enbart kovert eller enbart overt skönmålning var 0.78. Resultat efter korrektion simultant med båda skalorna för skönmålning var, för den koverta skalan 0.96 och för den overta 0.89. Överensstämmelsen var alltså god och det tycks som om enbart korrektion med den koverta skalan är fullt tillräcklig.

En omfattande översikt av forskning om skönmålning har publicerats av Uziel (Uziel, 2010). Traditionella skönmålningsskalor tycks fungera bättre än skalor för självbedrägeri, enligt hans översikt. Hans tes är emellertid att skönmålning inte kan mätas med skalor som mäter social önskvärdhet (skönmålningsskalor). Dessa skalor är nämligen enligt omfattande forskning korrelerade med en mängd andra variabler, mestadels med positivt innehåll. Därför skulle de mäta ett personlighetsdrag och kan inte vara mått på en "response style", enligt honom.

Det är en argumentation som är ganska vanlig och går tillbaka på Marlowe-Crownes tidigare arbeten (Crowne & Marlowe, 1964). Men logiken är svag. Även om skönmålningsskalor mäter tendensen att ge en överdrivet positiv bild av sig själv som en personlighetsegenskap kan de vara korrelerade med andra variabler. Det är ett absurt krav att de inte ska samvariera med något annat. I själva verket *måste* de samvariera med de personlighetsdrag som man önskar korrigeras för skönmålning, annars försvinner vitsen med hela proceduren. (Det blir ingen effekt av korrektionen). Det finns också som vi ska se positiva resultat som stödjer korrektionen för skönmålning (Sjöberg, 2009; Sjöberg & Wolgers, 2012).

Så låt gå för att det finns en rad samband mellan skönmålningsskalor och andra variabler. Hur stor kan effekten av sådana samband vara? Finns det en risk att man "överkorrigerar", dvs eliminerar varians som är av verklig betydelse och inte bara felet i mätningen, felet som beror på att den testade gett en överdrivet positiv bild?

Att det finns en korrelation mellan skönmålningsskalorna och substantiella personlighetsskalor har ibland lagts till grund för argumentet att man "korrigerar bort sann varians". En korrelation är emellertid just vad som bör förväntas om en del av variansen i personlighetsskalorna avspeglar skönmålning. Det är denna korrelation som gör det möjligt att genom korrektion avlägsna – åtminstone delvis – effekterna av skönmålning.

Det har visat sig att skalor av typ Marlowe-Crowne är känsliga för situationen som testningen görs i och vilka villkoren är. Detta är inte att vänta sig om skalan är ännu en personlighetsskala bland många andra. Ett citat:

*As a response validity scale intended as an index of the amount of positive distortion, the research base overwhelmingly suggests that IM<sup>2</sup> scales are sensitive to situational demands to respond in a desirable fashion". (Burns & Christiansen, 2006), p.11.*

McCrae och Costa (McCrae & Costa, 1983) publicerade en artikel som ofta åberopas som stöd för att skönmålningsskalor mäter "substans" snarare än tendensen att ge en överdrivet positiv bild av sig själv. Deras studie utnyttjade bedömningar av make/maka av personlighetsvariabler som kriterier. De fann att kontroll för skönmålning i stort sett inte påverkade korrelationen mellan testskalor och dessa kriterier. Vissa av testskalorna var korrelerade med skönmålning, och det var även vissa av kriterierna, i samma riktning. Data var emellertid insamlade i en forskningsstudie, inte i skarpt läge inför urval. Man kan inte utan vidare generalisera till skarpt läge. Effekterna av korrigerings på sambanden mellan

---

<sup>2</sup> Impression management

testskalor och kriterier var dessutom ytterst små. Kan man av sådana resultat dra slutsatser som att (a) skönmålning inte förekommer och (b) skönmålningsskalor i själva verket mäter ”personlighet” eller ”anpassning”? Det är onekligen en mycket långtgående tolkning. Mest intressant är emellertid följande punkt. Kriterierna var bedömningar av make/maka av vad de trodde att testpersonen hade svarat på testet, item för item, inte vad de anåg att han eller hon borde ha svarat för att vara sanningsenlig. En korrelation med skönmålningsskalan är därför högst rimlig, om man antar att maken/makan var medveten om att partnern gärna framställde sig själv som en socialt ”korrekt” person. Den möjligheten skulle vara intressant att pröva i fortsatt forskning där partnern också får besvara skönmålningsskalan på det sätt som han/hon tror att partnern gjort.

Uziels artikel är bra som exempel på argumentation mot korrektion för skönmålning. Det händer ju att man påstår att skönmålning inte förkommer och att den, om den skulle förekomma, saknar betydelse. Uziel lägger till påståendet att den inte går att mäta. Dessa argument är empiriskt felaktiga och även ologiska, rentav hyperkritiska<sup>3</sup>.

Tesen att skönmålningsskalor mäter ”personlighet” anses av många ha stöd i att det ofta finns korrelationer mellan sådana skalor och personlighetsvariabler. Men är det rimligt att tro att dessa skalor kan mäta många olika sorters personlighetsdimensioner? Tabell 3 senare i denna rapport visar att många personlighetsskalor hade höga samband med skönmålning, t ex dessa:

Utåtvändhet  
Följsamhet  
Emotionell stabilitet  
Noggrannhet  
Uthållighet  
Samarbetsvilja  
Positiv grundhållning  
Självförtroende  
Social förmåga  
Emotionell intelligens

Det verkar inte rimligt att en enda skala skulle kunna mäta det varierande innehållet i alla dessa testvariabler. Däremot är det rimligt att man får höga samband om de alla är känsliga för skönmålning.

## ***Validering av korrektion för skönmålning***

UPP-testet har använts vid testningar av sökande till officersutbildning (Sjöberg & Wolgers, 2012). Där kunde vi jämföra 4 olika grader av ”skarpt läge” med normdata med avseende på overt och kovert skönmålning. Betingelserna beskrivs i Tabell 2.

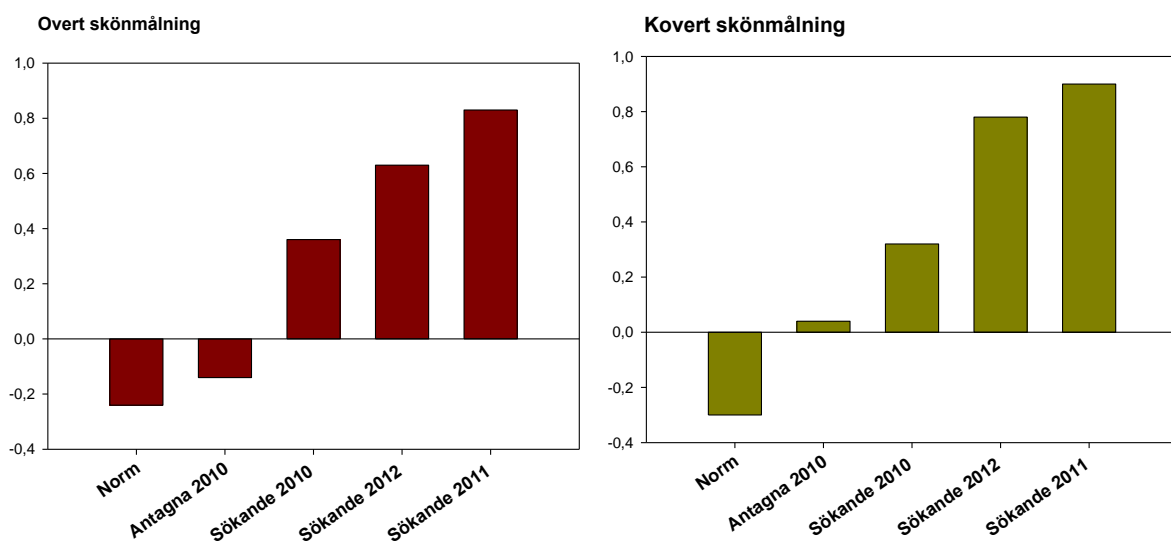
---

<sup>3</sup> Hyperkritiska diskurser av den här typen utmärks av att man söker med ljus och lykta efter något sätt att skjuta en teori eller metodik i sank, och i sin iver tappar man kontakten med logik, sansat omdöme och sunt förnuft.

**Tabell 2. Betingelser för testning, varierande grad av hur "skarp" testningen var**

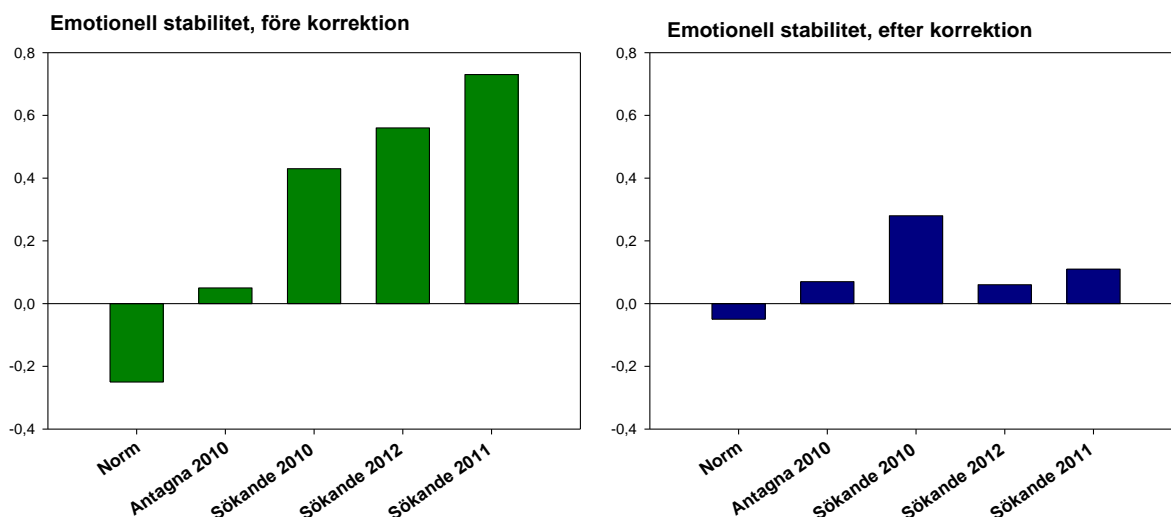
Grupp	Det uppfattades som att resultaten kunde påverka antagningsbeslutet	Det var klart utsagt att testningen var frivillig	Testningen var anonym	Testningen gjordes inom Försvarsmakten
Norm	Nej	Ja	Ja	Nej
Redan antagna	Nej	Ja	Nej	Ja
Sökande 2010	Möjligen	Ja	Nej	Ja
Sökande 2012	Möjligen	Ja	Nej	Ja
Sökande 2011	Ja	Nej	Nej	Ja

Se Fig. 2 för skillnader i skönmålning. Skillnaderna mellan grupperna var högt signifikanta och absolut sett mycket stora (>1 standardavvikelse). Det finns ingen anledning att tro att grupperna skilde sig i personlighetsdimensionen "Need for Approval".



**Figur 2. Overt och kovert skönmålning i fem grupper av testade. X-axeln: ökande grad av skarpt läge vid testning. Y-axeln: standardiserad skala.**

Fig. 3 visar variationen i testskalan emotionell stabilitet före och efter korrektion för skönmålning. Ca 95 % av variationen mellan grupperna försvann efter korrektion.

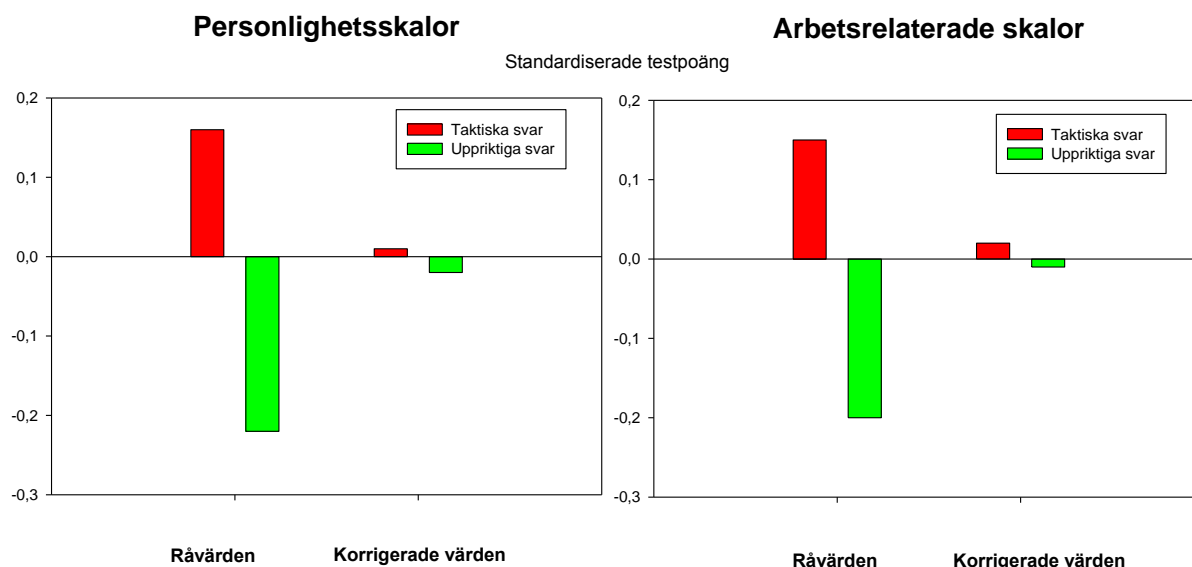


**Figur 3. Emotionell stabilitet i fem grupper av testade, före och efter korrektion för skönmålning. X-axeln: ökande grad av skarpt läge vid testning. Y-axeln: standardiserad skala.**

Fig. 4 visar resultatet av ett experiment som beskrivits närmare i (Sjöberg, 2008). 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, men ändå trovärdig. Det kanske innebär att du bluffar en del, men det är just det som vi har för avsikt att studera i den här undersökningen. Känn dig därför fri att svara taktiskt!”

Skillnaden var stor före korrektion men försvann till 90-95 % efter korrektion, se Fig. 4.



**Figur 4. Genomsnittliga effekter av korrektion för skönmålning, standardiserade skalor.**



Skalorna för skönmålning i *UPP*-testet har också validerats på följande sätt. Testningen avslutades med att testpersonen ombads ta ställning till följande påstående:

”Jag besvarade testfrågorna så att mina svar skulle ge en positiv bild av mig”

Detta påstående skulle bedömas genom att välja något av följande svarsalternativ:

- Nej, absolut inte
- Nej, i stort sett inte
- Tveksamt
- Ja, i stort sett
- Ja, absolut

Fördelningen av svar på denna fråga var intressant, se Tabell 3. Ca 30 % medgav skönmålning, eller förnekade det inte. Det stämmer bra med den meta-analytiska uppskattningen att ca 30 % skönmålar (Griffith & Converse, 2012), G33 och är intressant utifrån Zickars och Sliters (Zickar & Sliter, 2012)formulering att

*”very few people would admit to faking in actual settings”* (p. 124)

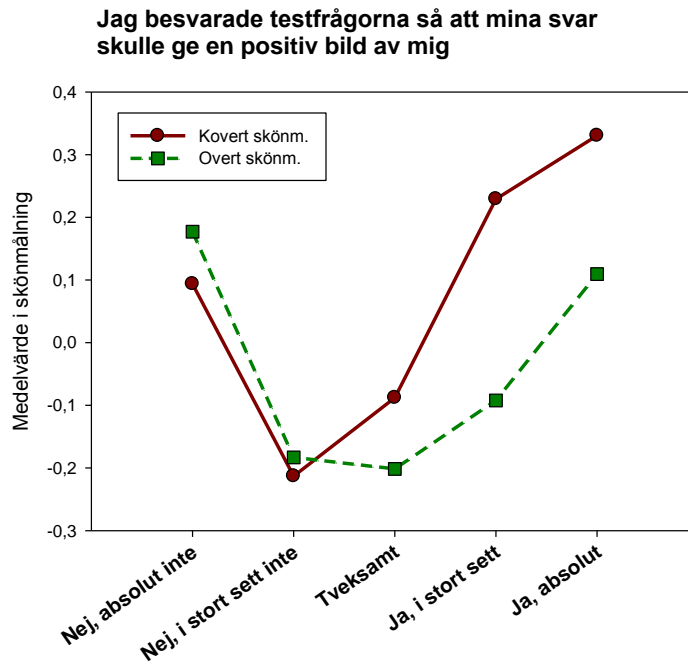
Ytterst få tycks ha prövat om detta är korrekt, såvitt jag vet ingen annan, men det verkar ju suntsförnuftsmässigt att tro att så skulle vara fallet. Trots det tycks det alltså som om det går att få en del information om vilka som har skönmålat genom att direkt fråga om det.

Tabell 3. Svar på påståendet ” Jag besvarade testfrågorna så att mina svar skulle ge en positiv bild av mig”

	Frekvens	Procent av alla data	Procent av avgivna svar
1 Nej, absolut inte	459	42,1	42,8
2 Nej, i stort sett inte	291	26,7	27,1
3 Tveksamt	207	19,0	19,3
4 Ja, i stort sett	88	8,1	8,2
5 Ja, absolut	27	2,5	2,5

Hur samvarierade svaren på påståendet med de två skalorna för skönmålning (overt och kovert)? Se Fig. 5<sup>4</sup>.

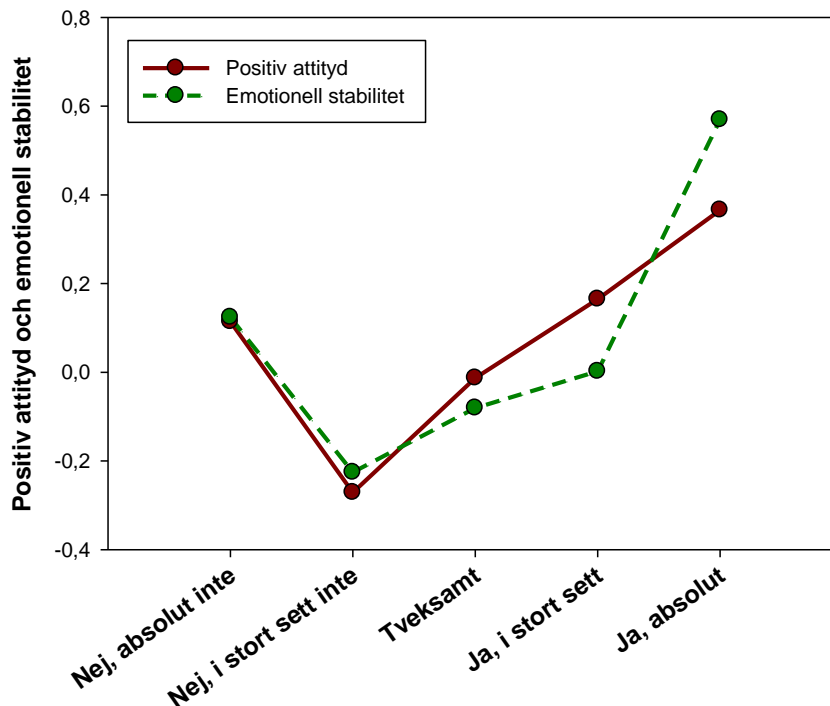
<sup>4</sup> Data från 484 testningar med sökande till tjänster som chefer eller specialister, huvudsakligen i det privata näringslivet.



**Figur 5. Skönmålning i relation till svar på frågan om den testade försökt ge en positiv bild av sig själv.**

Sambandet mellan svaren på påståendet och skönmålning var högt signifikant. Det är slående att overt och kovert skönmålning ger mycket likartade resultat och att höga värden främst erhöles för dem som förnekande skönmålning *eller* medgav det.

Vilka personlighetsvärden gav de testade i sina självrapporter? Se Fig. 6 för genomsnittliga värden av Positiv attityd och Emotionell stabilitet.



**Figur 6. Positiv attityd och emotionell stabilitet i relation till svar på frågan om den testade försökt ge en positiv bild av sig själv.**

Figuren visar att de som förnekande *eller* medgav skönmålning fick de högsta värdena i dessa två viktiga (och positiva) personlighetsvariabler. Att svara ”Nej, absolut inte” på frågan om man skönmålat är ju också ett uttryck för skönmålning, särskilt i skarpt läge, som här var fallet.

### ***Personlighet eller skönmålning?***

Det är vanligt att man sätter ett likhetstecken mellan skönmålning som den definieras av en skönmålningsskala och ”need for approval” – slutsatsen blir att skalan inte alls mäter tendensen att försöka få ett bra resultat på testet genom att ge vad man tror är önskade svar (”faking”) utan ”personlighet (Crowne & Marlowe, 1964). Personlighet är något som är generellt över tid och rum, inte så specifikt som att ge distorderade svar på ett visst test vid ett visst tillfälle, så tycks argumentet vara tänkt. Olika tolkningar är emellertid möjliga.

Marlowe-Crownes skala för social önskvärdhet var inte ursprungligen tänkt som en personlighetsskala utan som en skala på ”faking”. Författarnas eget arbete ledde dem senare till slutsatsen att det snarare handlade om personlighet. Men vad ska vi tro om den saken? En aktuell avhandling av Barger (Barger, 2012) ger en mycket användbar översikt av mätning av ”Need for Approval”, inklusive fullständiga förteckningar av de item som ingår i skalorna. Förutom Marlowe-Crownes skala studerade han DAS-skalan (Weissman & Beck, 1978) samt GABS – skalan (Bernard, 1998; Lindner, Kirkby, Wertheim, & Birch, 1999)<sup>5</sup>. DAS och

<sup>5</sup> Marlowe-Crownes skala, DAS-skalan och GABS återges i Appendix 1.

GABS är konstruerade för att mäta "Need for Approval" och har validerats i skilda sammanhang. GABS och DAS korrelerade 0.63 i Bargers studie, men Marlowe-Crownes skala korrelerade bara -0.08 och -0.15 med respektive DAS och GABS. Det verkar alltså inte som om Marlowe-Crownes skala mäter den postulerade personlighetsdimensionen. Det är tydligt från skalans konkreta innehåll att DAS är mycket konsekvent inriktad just på behovet att vara omtyckt medan Marlowe-Crownes skala bygger på en samling av olika handlingar som är vanliga, men anses icke önskvärda, eller ovanliga, men anses önskvärda.

Betrakta lite närmare de typiska påståendena i Marlowe-Crownes skala, t ex "I sometimes think when people have a misfortune they only got what they deserved". Om man svarar "nej" på denna fråga bidrar det positivt till poängen i skalan. Samtidigt är det troligen ett ganska vanligt beteende. En möjlig tolkning av varför någon svarar nej är att han eller hon vill framstå som en sympatisk person, och att detta är ett uttryck för en generell sådan tendens. En annan tolkning är önsknigen att framstå som en sympatisk person *i testningen*. *Inget fall är det bevisat att det fråga om ett generellt och tydligt behov av att vara omtyckt*. Det är en tolkning som går mycket längre, och som enligt vad vi sett inte får stöd när man relaterar Marlowe-Crownes skala till övriga skalor som mera direkt mäter behovet av att vara omtyckt. Marlowe-Crownes skala som mått på "Need for Approval" är ett exempel på så kallade "subtle items", dvs item som inte tydligt ger uttryck för det begrepp som ska mätas, utan där det krävs en mer eller mindre långsökt tolkning. Forskningen om sådana item har inte visat att de var användbara för att minska inflytandet av skönmålning (Weed, Ben-Porath, & Butcher, 1990).

Det kan tilläggas att Marlowe-Crownes skala i en meta-analys (Ones, Viswesvaran, & Reiss, 1996) befanns ha mycket svaga samband med arbetsresultat, omkring 0. Det skulle knappast ha varit fallet om den skalan mäter personlighet. Andra studier har visat tydliga negativa samband mellan skönmålning och arbetsresultat (Peterson & Griffith, 2006). Kortsiktiga positiva bedömningar kan en anställd uppnå i jobbet genom olika taktiska sätt att bete sig men ganska snart genomskådas han eller hon (Higgins, 2003 #10937). Idén att den som kan fejka testresultatet och också är motiverad att fejka ett bra resultat på jobbet är märklig. För är det inte bara en fråga om motivation om man lyckas på jobbet, utan också kompetens. Kan man fejka kompetens?

Slutsatsen blir att Marlowe-Crownes skala inte mäter personlighet, speciellt inte "Need for Approval", utan - som ursprungligen avsett - tendensen att svara socialt önskvärt i en testsituation. Detta hindrar inte att en sådan tendens kan återfinns i flera situationer, t ex i olika test eller vid olika tidpunkter. Man har funnit för gruppen av testpersoner med höga skönmålningsvärden att testvaliditeten är låg (White, 2001 #10931), något som är att förvänta om skönmålningsskalorna faktiskt mäter snedvridning av testresultaten. Det finns, som vi ska se i nästa avsnitt, samband mellan Marlowe-Crownes skala och svarstid, vilket också stämmer med antagandet att Marlowe-Crownes skala mäter skönmålning.

## Svarstid

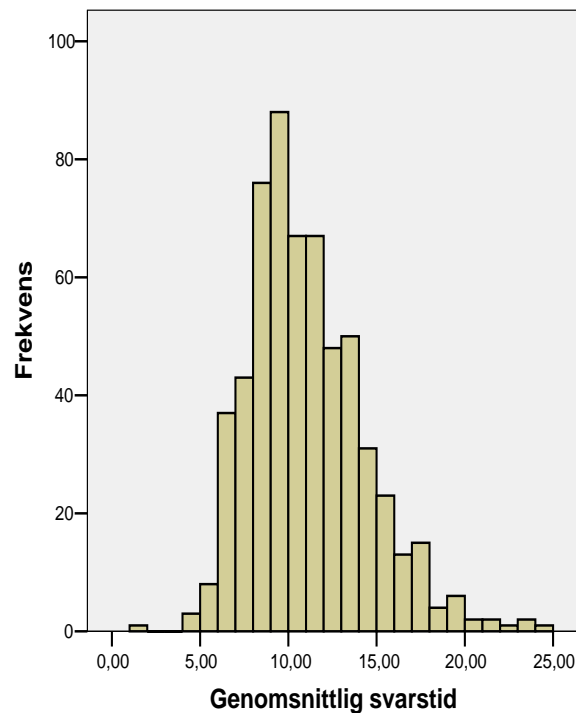
Flera forskare har föreslagit, och funnit stöd för, antagandet att längre svarstider är förknippade med skönmålning (Holden, 1998; Holden & Hibbs, 1995; Holden & Kroner, 1992; van Hoof & Born, 2012). I UPP-testet registreras alla svarstider i sekunder<sup>6</sup>. Svarstiden beror bl a på antalet ord i en testuppgift. Korrelationen mellan genomsnittlig svarstid och

---

<sup>6</sup> Tider >45 sekunder har uteslutits ur följande analyser då det är troligt att de är förknippade med avbruten testning.

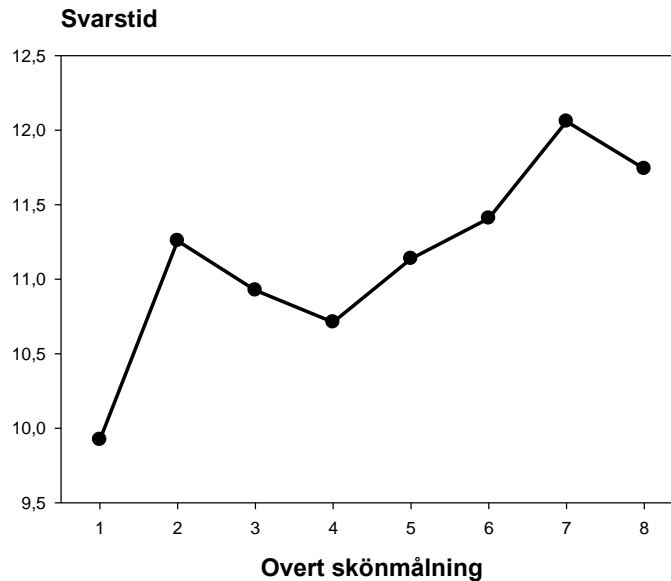
antalet ord var 0.57,  $p < 0.0005$ . Variationen i genomsnittlig svarstid per item var 6.1 - 20.0 sekunder.

I den aktuella datamängden var den genomsnittliga svarstiden per person 11.1 sekunder, beräknat över samtliga personlighetsfrågor. Fördelningen framgår av Fig.7. Sambandet mellan svarstid och overt skönmålning framgår av Fig. 8.



**Figur 7. Fördelning av den genomsnittliga svarstiden.**

Envägs variansanalys med overt skönmålning som beroende variabel och grupperingen i genomsnittlig svarstid gav  $F(4, 582) = 2.615$ ,  $p = 0.034$ , alltså signifikant. För kovert skönmålning fanns inget tydligt samband med svarstid även om en svag trend av samma typ som för overt skönmålning kunde skönjas. Fig. 8 visar sambandet svarstid - skönmålning efter en gruppering av testpersonerna. Högre skönmålning medförde en tydlig höjning av svarstiden.



**Figur 8. Sambandet mellan genomsnittlig svarstid i sekunder (per item) och overt skönmålning i 8 grupper av testade personer**

Resultaten av dessa analyser kan sammanfattas på följande sätt:

1. Det fanns ett starkt samband mellan antal ord i ett item och den genomsnittliga svarstiden
2. Det fanns ett något svagare men ändå tydligt samband mellan svarstid och skönmålning

Punkt 1 visar att data gett resultat som innebär en validering av mätningen av svarstid. Punkt 2 är en validering av mätningen av overt skönmålning.

En intressant aspekt på svarstid är den eventuella tidsbegränsningen som gäller för att ta testet. Sådan tidsbegränsning är troligen ganska ovanlig. Det har i en studie visat sig att tidsbegränsning minskar tendensen till skönmålning, dock enbart hos testade med låg kognitiv förmåga (Komar, Komar, Robie, & Taggar, 2010).

## Hur vanlig och stark är skönmålning?

Genomslaget av skönmålning är olika stort för olika testskalor, se Tabell 4 för data från testningar i skarpt och oskarpt läge med *UPP*.

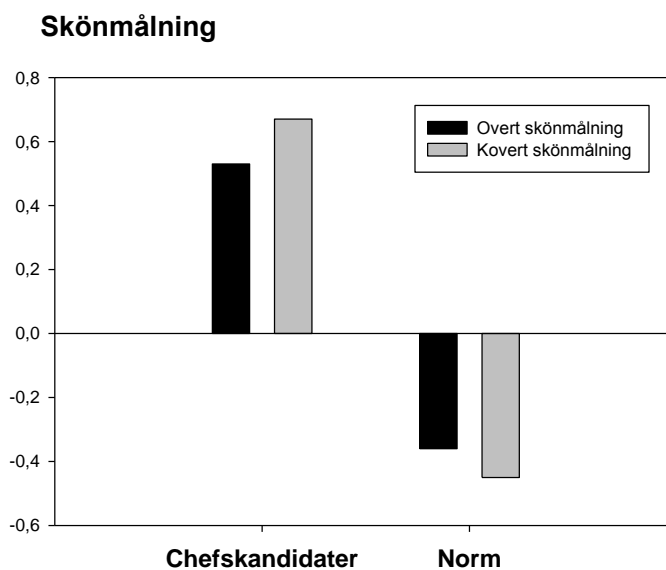
Tabell 4. Variansandelar förklarade av skönmålning hos sökande och redan anställda.

Skala	Av skönmålning förklarad variansandel i gruppen av sökande (skarpt läge)	Av skönmålning förklarad variansandel i gruppen av anställda (oskarpt läge)
Utåtvändhet	0.242	0.135
Följsamhet	0.321	0.172
Emotionell stabilitet	0.378	0.268
Öppenhet	0.158	0.067
Noggrannhet	0.421	0.255
Uthållighet	0.403	0.315
Samarbetsvilja	0.502	0.384
Positiv grundhållning	0.306	0.351
Självförtroende	0.374	0.289
Social förmåga	0.245	0.134
Emotionell intelligens, självrapport	0.383	0.314
Emotionell intelligens, emotionsbedömningar	0.041	0.019
Kreativitet	0.006	0.000
Perfektionism	0.042	0.015
Narcissism	0.010	0.041
Social säkerhet	0.182	-
Arbetsstillfredsställelse	0.039	0.153
Arbetsvilja	0.104	0.246
Resultatorientering	0.158	0.112
Förändringsvilja	0.237	0.120
Arbetsintresse	0.053	0.039
Balans arbete – övrigt liv	0.211	0.123
Skönmålning 1 (kovert)	-	-
Skönmålning 2 (overt)	-	-
Hedonisk ton	0.327	0.257
Aktivitet	0.208	0.110
Spänning	0.042	0.082
Humör, sammanslaget	0.296	0.237
Attityd till testet	0.004	0.022
Jagstyrka	0.538	0.498
Chefspotential	0.451	0.442
Stresstålighet	0.543	0.574
Median (exkl. aggregerade variabler)	0.211	0.135

Tabellen visar betydande skillnader mellan testskalorna. Den visar också ett mycket större genomsnitt för skönmålning i gruppen av sökande än i gruppen av redan antagna, ca 50 % större i genomsnitt.

Jämförelser mellan jobbsökande ("applicants") och redan anställda ("incumbents") brukar visa på betydande skillnader på personlighetstest (Donovan, et al., 2003; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990). I en grupp av sökande brukar validiteten vara lägre än bland redan anställda eller andra som tar testet i oskarpt läge (Hough, 1998). Detta kan bero på att skönmålning har ett större genomslag i en sådan grupp än i en grupp som tagit testet i oskarpt läge. Se Tabell 4 för exempel på sådana resultat. Ännu större skillnader brukar det bli mellan en grupp som svarar "ärligt" och en som uppmanas att ge en maximalt positiv bild av sig själva (Tett, et al., 2006). Litteraturen på området är emellertid inte entydig. Somliga – en minoritet bland forskarna – säger att skillnaderna är små, andra finner att de är betydande. I skarpt läge har man goda skäl att tro att skönmålningen är av betydande storlek. Praktiskt arbetande rekryteringskonsulter har ofta funnit att test av självrapporttyp som inte korrigerar för skönmålning är av tveksamt värde eftersom så många testade har mycket höga värden. Ibland visar det sig också att kvinnliga jobbsökande missgynnas, liksom invandrare (Sjöberg, 2010), eftersom de tycks skönmåla mindre (Sjöberg, 2005). Yngre personer och män skönmålar mera än äldre och kvinnor enligt resultat av Ziegler som citeras av Heggstad (Heggstad, 2012).

Data från användning av UPP utvärderades i skönmålning för en stor grupp av chefskandidater, testade i skarpt läge av rekryteringskonsulter, och en normgrupp (N = 1080) som testades i oskarpt läge (N = 726) . Se Fig. 8.



**Figur 9. Skönmålning hos chefskandidater (skarpt läge) och norm (oskarpt läge).**

Figuren visar att skönmålning är i genomsnitt mycket starkare hos jobbsökande än hos testade i oskarpt läge, ca 1 standardavvikelse.

I en meta-analys av prevalensen hos skönmålning hos jobbsökande fann Griffith och Converse (Griffith & Converse, 2012a) att den låg i genomsnitt på 30 % - se deras Tabell 3.1 p.48.

Bluffning är vanligt också i andra delar av den information som en sökande lämnar vid ansökan eller intervju (Hough, 1998). Förskönade meritförteckningar tycks vara vanliga och bli allt vanligare. På Internet hittar man många tips om hur man ska upptäcka bluffen.



Levashina och Campion (Levashina & Campion, 2007) fann att skönmålning var vanlig i intervjuer. De har utvecklat en skala för att mäta "faking" i intervjuer. Detta förefaller vara en idé väl värd följa upp i ny forskning.

## **Vad är det som ger upphov till skönmålning?**

Det troligaste svaret på rubrikens fråga är att vissa testpersoner skönmålar om de är angelägna om att få ett bra resultat och tror att de kan åstadkomma det genom att ge taktiska svar på testet. Detta har visat sig stämma i många studier. Man kan emellertid fråga sig om inte också andra faktorer är av betydelse, se en översikt av Ellingson (Ellingson, 2012), som går igenom litteraturen och behandlar ett antal faktorer, men sambanden var genomgående låga. Tonković fanns samband mellan skönmålning och risktagande och låg ärlighet (Tonković, 2012).

## **Skönmålning och teststruktur**

Skönmålning kan ha den effekten att faktorstrukturen hos det använda testet förändras och kommer att av spegla en övergripande "halofaktor" (Schmit & Ryan, 1993; Zickar & Robie, 1999). Detta är emellertid inte alltid fallet. *UPP*-testets struktur har jämförts i skarpt och oskarpt läge. Resultaten visade mycket likartade strukturer, trots att nivån på skönmålning av dramatiskt olika (Sjöberg & Wolgers, 2012). Andra forskare har fått liknande resultat med andra test (Bradley & Hauenstein, 2006).

## **Metodik för att handskas med skönmålning**

### ***Suppressormodellen för korrektion för skönmålning***

Den enkla korrelationen mellan skönmålning och personlighetsskalor åberopas ibland som tecken på att dessa skalor mäter substantiella personlighetsdrag. Det är ett förvånande argument eftersom en korrelation är precis vad som behövs om man ska lyckas med att använda skönmålningsskalor för korrektion. Skönmålningsskalan kan nämligen som avsett fungera som en "suppressor" för sambandet mellan personlighet och ett kriterium (t ex arbetsprestation) om två villkor är uppfyllda beträffande denna skala

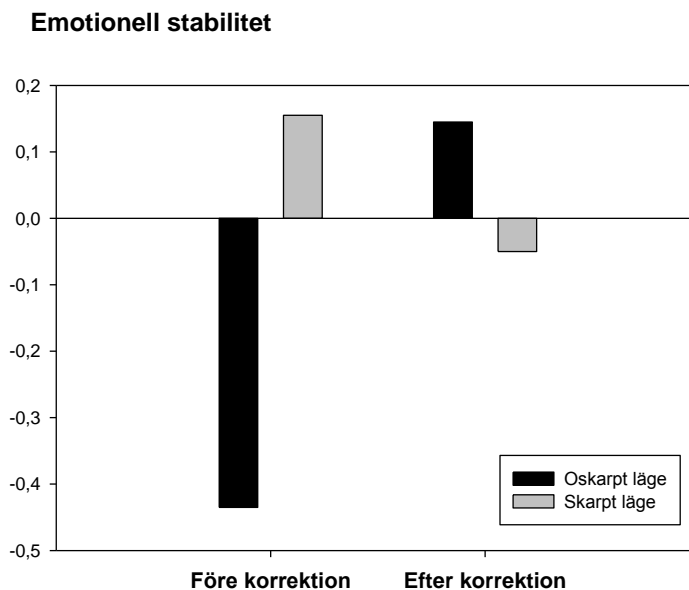
1. Den har negativ eller ingen korrelation med kriteriet
2. Den har en betydande positiv korrelation med personlighetsskalan.

Båda villkoren är normalt uppfyllda, även om sambanden oftast inte är starka nog för att ge stora effekter på validiteten hos personlighetsskalan. För vissa kriterier är det första villkoret vanligen inte uppfyllt; man finner i stället positiva samband med skönmålning. Då leder korrektionen enligt suppressormodellen till minskad validitet (i allmänhet också det en svag effekt).

### ***Normativ metodik***

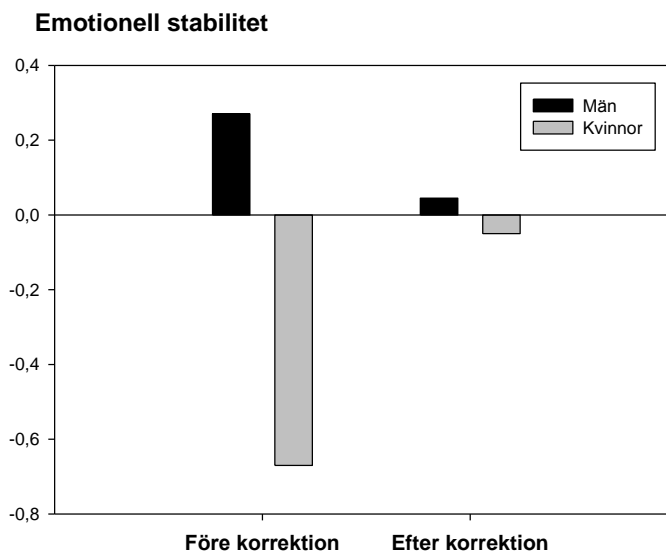
Att mäta separat, med en eller flera skalor, den tendens de svarande har att skönmåla har visat sig fungera bra. Man kan korrigera bort effekterna av skönmålningen, som ju kan ha stora effekter i enskilda fall (Meehl & Hathaway, 1946). Ett exempel kan ges från en studie av sökande till Handelshögskolan i Stockholm, jämförda med en grupp som tog testerna anonymt

(Engelberg & Sjöberg, 2005). Testningen inför antagningen var en typisk ”skarp” testning där mycket stod på spel för de testade. Figur 10 visar resultatet av korrektion av personlighetsdimensionen emotionell stabilitet.



**Figur10. Effekten av korrektion, jämförelse mellan testning i skarpt läge och under anonymitet. Skala 1-9, Medelvärde=5, Standardavvikelse=2.**

Observera att en stor skillnad till förmån för de testade i skarpt läge efter korrektion försvann och t o m ändrades till en liten fördel för dem som testades helt anonymt. Hur viktigt det kan vara med korrektion visas i Fig. 9 som ger könsskillnader före och efter korrektion i en grupp av personer som var kandidater till chefsjobb.

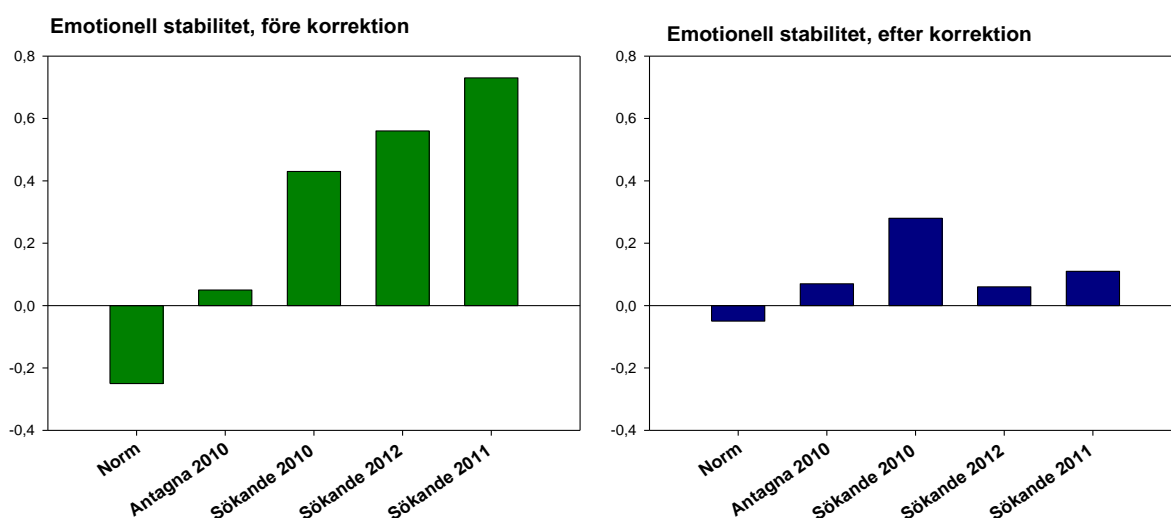


**Figur 10. Könsskillnader i emotionell stabilitet före och efter korrektion. Skala 1-9, Medelvärde=5, Standardavvikelse=2.**

Fig. 9 visar en mycket stor fördel för männen före korrektion för skönmålning. Efter korrektion försvann den nästan helt. Resultatet avspeglar att kvinnorna hade en mycket svagare tendens att ge en överdrivet positiv bild av sig själva än vad männen hade.

Situationen är likartad när det gäller testning av invandrare. De har visat sig skönmåla något mindre än personer av svensk härkomst och missgynnas därför av test som inte korrigerar för skönmålning (Sjöberg, 2010).

I en studie av sökande till officersutbildning användes samma metodik (Sjöberg & Wolgers, 2012). Samtliga gruppers resultat jämfördes före och efter korrektion, se Fig. 11 för testvariabeln Emotionell Stabilitet. Drygt 95 % av effekten av skönmålning eliminerades. Mycket stora och starkt signifikanta gruppskillnader före korrektion försvann och var inte längre statistiskt signifikanta.



**Figur 11. Emotionell stabilitet i fem grupper av testade, före och efter korrektion för skönmålning. X-axeln: ökande grad av skarpt läge vid testning, Y-axeln: standardpoäng (z-skala).**

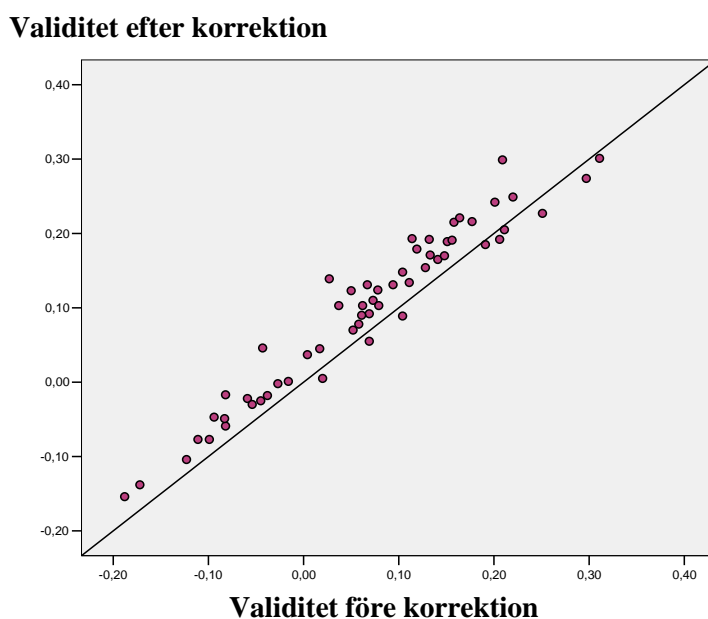
Forskningen om skönmålning svar på personlighetstest har i allmänhet använt sig av en experimentell uppläggning. Det är ganska vanligt att man mäter skönmålning som skillnaden mellan två betingelser, nämligen ärliga svar och skönmålning svar. Forskningen har inte gett tydligt stöd för denna metodik (Sackett & Lievens, 2008). Skälet är troligen att skönmålning i en reell antagningstestning inte är samma process som skönmålning i ett experiment där man uppmanas att ge en i viss mån förvanskad bild av sig själv. Märk också att detta mått på skönmålning är svårt eller omöjligt att använda i tillämpat arbete där man vill mäta skönmålning hos varje testad person. Däremot är det helt klart från forskningen att skönmålningsskalor kan skilja vilka som medvetet bluffat efter att ha blivit uppmanade till det, från andra som tagit ett test under vanliga instruktioner att svara ärligt (Visweswaran & Ones, 1999). Fernando och Chico (Fernando & Chico, 2001) fann dessutom att skönmålningsskalor var betydligt bättre på att upptäcka bluffare än mått som baserades på en alternativ idé, nämligen hur väl en persons testsvar kunde förklaras med en matematisk modell för testet (Visweswaran & Ones, 1999).

Det finns en del forskning som tycks visa att korrektion av den typ jag förespråkar inte leder till högre validitet hos testerna, men inte heller till lägre (Ones & Visweswaran, 1998)<sup>7</sup>. Spelar

<sup>7</sup> Deras slutsatser är osäkra eftersom deras data till övervägande del byggde på testningar med anställda, inte med sökande. Korrelationsstrukturen tenderar att vara olika för dessa två grupper.

det då någon roll om man korrigerar eller inte? Svaret är att det gör det, och skälet är att det i enskilda fall kan vara mycket betydelsefullt om man korrigerar. Tätgruppen av sökande till ett jobb eller en utbildning består, om man inte korrigerar, regelmässigt till stor del av dem som svarar i hög grad taktiskt och inte uppriktigt.

Härtill kommer att aktuell forskning visat att korrektion faktiskt kan ge högre validitet, se Fig. 12 som bygger på testning av 166 chefer på mellannivå inom Stockholms Läns Landsting med *UPP*-testet (Sjöberg, Bergman, Lornudd, & Sandahl, 2011). Kriterierna var 360-gradersbedömningar av ledarskapets effektivitet som gjordes av deras medarbetare, egna chefer och kolleger (Arvonen, 2002). Figuren ger endast data för medarbetarnas bedömningar men resultatet var likartat för de övriga bedömningsserierna.



**Figur 12. Validitet hos skalor i *UPP*-testet efter korrektion för skönmålning prickade mot motsvarande data före korrektion.**

Den korrektion med hjälp av skalor för mätning av skönmålning som görs i *UPP*-testet har en stor fördel framför andra metoder: den är specifik för varje testvariabel. Det visar sig nämligen att sambandet mellan skönmålning och testvariablerna är av varierande styrka, se Tabell 4. Man kan inte göra en global bedömning av om en person är trovärdig. Det är något som varierar beroende på vilken personlighetsegenskap man diskuterar, se Tabell 3.

Normativa format medför vissa problem utöver skönmålning. Det har hävdats att normativa svarsformat kan medföra vissa besvärliga felkällor, nämligen att personer varierar i tendens att hålla med om påståenden oavsett innehåll (jasägartendens) och vad gäller den intra-individuella variationen i svar (Brown & Bartram, 2009). Somliga utnyttjar skalan fullt ut, andra har en tendens att använda bara några få svarskategorier. Dessa svarstendensers samband med testskalorna reduceras emellertid starkt efter korrektionen för skönmålning (Sjöberg, 2009).

Ett annat möjligt problem är att de korrigerade skalorna har låg reliabilitet<sup>8</sup>, då skulle det t ex vara svårt att få skillnader mellan grupper som tagit testet i olika grader av skarpt läge. Skälet till denna misstanke kunde vara att det är ganska höga samband mellan skönmålning och de okorrigerade testskalorna. Om man korrigerar kan det ses som att man eliminerar sann varians i personlighet. Är det verkligen den effekt som uppstår? Frågan kan besvaras genom att studera reliabiliteten hos skalorna före och efter korrektion. I normdata analyserades samtliga testvariablers reliabilitet före och efter korrektion.

Det visade som väntat att det blev en viss sänkning av reliabiliteten efter korrektion 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 korrigera för skönmålning:

- Testvärdena ger efter korrektion en sannare bild av testpersonen eftersom eventuella försök att svara taktiskt i stort sett inte påverkar dem
- Testvärdena efter korrektion är mindre påverkade av slumpfel i mätningarna i betydelsen felspridning, medan reliabiliteten minskade något

### ***Ipsativa svarsformat***

Ett försök att motverka effekterna av skönmålning är att gå över till ett format där de testade gör jämförande bedömningar, ett s.k. ipsativt format. Det ipsativa formatet är troligen det vanligaste försöket att lösa problemet med skönmålning. Vanliga exempel på detta förfarande ges i test som OPQ, Thomas, MPA och MBTI (Mabon, 2004). Här består valalternativen av uppgifter eller påståenden som är matchade i social önskvärdhet och man hoppas att valet av svarsalternativ därför inte påverkas av tendensen att svara taktiskt. Det visar sig emellertid att detta är en förhoppning som inte uppfylls. Det finns ett flertal problem med ipsativa testformat.

Ett första problem är att det är svårt att åstadkomma matchning i social önskvärdhet och att denna dessutom inte är giltig från en person till en annan. Vidare är statistiska analyser av data från denna typ av test svåra eller omöjliga att göra. Reliabiliteten kan inte bestämmas med vanlig itemanalys<sup>9</sup>. Det går inte heller att studera faktorstrukturen i en uppsättning ipsativa skalor.

När man använder ipsativa format ger man den testade en uppgift som kan vara mycket svår. Om man ska välja mellan två alternativ men inte gillar något av dem – hur gör man då? Ipsativa format är inte populära bland de testade (Harland, 2003). De tycker inte om att "jämföra päron och äpplen". Bowen, Hunter och Hunt (Bowen, Martin, & Hunt, 2002) jämförde de testades reaktioner på ipsativt och normativt format av OPQ. De skriver:

---

<sup>8</sup> Detta kan misstänkas eftersom de korrigerade variablerna har formen av differenser (residualer). Det har ofta påpekats i den psykometriska litteraturen att differenser har lägre reliabilitet än de variabler som används för att bilda differenserna. Farhågorna tycks dock vara överdrivna; effekterna behöver inte vara dramatiskt stora (Rogosa & Willett, 1983). Se även den tekniska manualen för UPP-testet (Sjöberg, 2011).

<sup>9</sup> Se Magnusson (Magnusson, 2003) för en introduktion till den klassiska testteorin.

*“Normative scales received more positive reactions from the participants on all items. Participants felt the normative format was easier to answer, allowed them to present themselves as they wanted to, was less confusing, and was more interesting.” (Sid. 249).*

Dessutom gick det fortare att besvara testet i ett normativt format. Det ipsativa formatet har inga kända fördelar framför det vanliga normativa. Personlighetstester används ofta som beslutsunderlag, där personer ska jämföras med varandra. Men ipsativa testpoäng är *definitionsmissigt* omöjliga att jämföra mellan personer. En sådan jämförelse är därför meningslös, men tyvärr verkar det som om en del testanvändare inte förstått detta. Modellanalys (Brown & Maydeu-Olivares, 2011) kan inte återskapa den information som saknas i svaren.

Det finns emellertid ett ännu värre problem. Om en person ligger högt i många dimensioner, men lite mindre högt i t ex emotionell stabilitet, kommer testet inte att upptäcka detta, utan han eller hon får ett lågt värde i emotionell stabilitet, *helt enkelt för att alla värden är relativa till individen själv*. Den emotionella stabiliteten kan vara utmärkt, men andra egenskaper ligger ännu högre och därför hamnar den långt ner på skalan - mycket vilseledande. En pedagogisk metafor är denna: antag att en 10-kampare kommer 1:a i 9 grenar men 2:a i höjdhopp. Ipsativt sett är han en dålig höjdhoppare! En person som ligger högt i alla eller flesta avseenden kan inte utvärderas med ett ipsativt test som OPQ.

Hur är det i praktiken? Är ipsativa test mera valida och mindre känsliga för skönmålning? Detta har länge förnekats (Hicks, 1970) men i senare arbeten av företrädare för OPQ-testet har det påståtts att så trots allt är fallet (Bowen, et al., 2002). Uppläggningsen i de citerade studierna var experimentell. Det var alltså ingen reell antagningstestning som Bowen et al. studerade. De jämförde ett ipsativt format med ett *okorrigerat* normativt format. Ingen torde vänta sig annat än att det senare formatet är känsligt för ”faking”. Poängen är ju att man kan och bör korrigera för skönmålning. I ett aktuellt arbete fann man att ett ipsativt format inte löste problemet med skönmålning svar (Heggstad, Morrison, Reeve, & McCloy, 2006).

I den psykometriska litteraturen är man numera tämligen enig om att ipsativa format dels inte löser problemet med ”faking”, dels medför stora egna problem som inte finns i normativa format (Converse, et al., 2006). Så skriver t.ex. Meade (Meade, 2004):

*”In sum, the standards required of tests used for employment selection are quite strict with regards to validity and reliability of the selection instruments- As such, the limitations inherent with ipsative measures pose too great a threat to the validity of the selection tools to make it a useful instrument for selection on a trait-by-trait basis.”*

---

Man bör alltså undvika test som använder ipsativa format. Det är möjligt att problemen är mindre utpräglade om man har många skalor eller valalternativ (Bowen, et al., 2002), men de kvarstår i princip och inga fördelar finns med det ipsativa formatet. Troligen har de en gång konstruerats i tron att man på detta sätt undviker ”faking”. Så är alltså inte fallet.

Eventuellt har test med ipsativa skalor högre validitet än normativa test (Jackson, Wroblewski, & Ashton, 2000). Till en del kan denna ökning av validiteten bero på att ipsativa test korrelerar med intelligens (Vasilopoulos, Cucina, Dyomina, Morewitz, & Reilly, 2006), vilket normativa test inte gör i samma utsträckning. Dilchert och Ones (Dilchert & Ones, 2012) skriver, i den mesta aktuella uppsummeringen av frågan om ipsativa format:

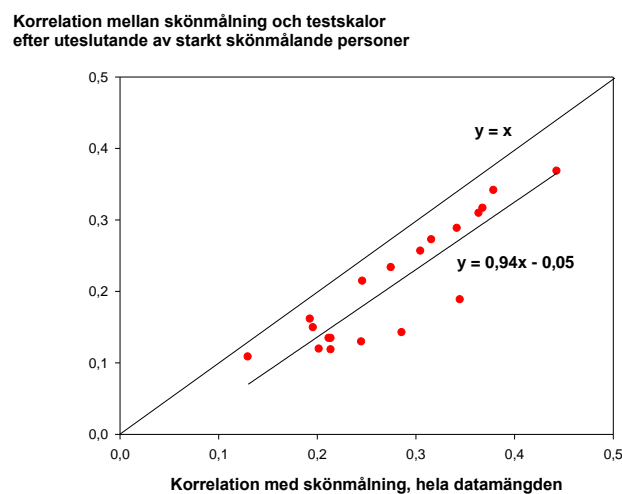
”Finally, and perhaps most importantly in terms of applied implications, forced-choice formats that result in ipsative scores are not a good means of preventing faking because they do not result in higher criterion-related validity.” (p.192)

Man uppnår alltså inte den fördel man eftersträvade med det ipsativa formatet och man får en mängd andra allvarliga problem på köpet. Dessutom tycks ipsativa format skapa starkare samband mellan personlighet, mätt med ipsativa test, och IK, vilket inte är önskvärt<sup>10</sup> (Christiansen, Burns, & Montgomery, 2005). Ipsativa format har visat sig vara en återvändsgränd i testmetodik. Ju förr de överges desto bättre.

## Användning av ett gränsvärde

Det förekommer att man försöker gardera sig mot effekterna av skönmålning genom att utesluta de testpersoner som ligger högt i det avseendet, t ex de 25 % som ligger högst. Vi ska nu skaffa oss en uppfattning om effekterna av att använda ett sådant gränsvärde i en skala för skönmålning. Får man bort samvariationen mellan testresultat och skönmålning på det sättet, i gruppen som består av de återstående 75 procenten? Följande analyser gjordes i UPP-testets normdata för att svara på den frågan.

(1). Korrelationer beräknades mellan alla testets variabler och overt skönmålning för hela gruppen. Därefter uteslöts de 25 % av de testade som hade högst värden i overt skönmålning och korrelationer beräknades i gruppen som bestod av de 75 % återstående. De två serierna av korrelationer har prickats i Fig. 13.



**Figur 13. Korrelation mellan testvariabler och skönmålning för hela datamängden och efter uteslutande av de 25 % som hade starkast tendens till skönmålning.**

Figuren visar att de två serierna av korrelationer var starkt relaterade. Mer intressant är att korrelationerna sjönk efter det att extremgruppen i skönmålning hade uteslutits, men ingalunda till noll. Tillämpar man regressionsekvationen i figuren på en typisk korrelation på 0.30 i hela gruppen får man värdet 0.23 efter uteslutande av 25 % av de testade. I termer av förklarad varians sjönk denna alltså från 9 % till 5 %. Enligt denna grova uppskattning är

<sup>10</sup> Man vill ha personlighetstest med låg korrelation med IK, för att på det sättet säkerställa att man får ett tillskott i information utöver IK.

ungefär hälften av effekten av skönmålning kvar, även om man rensar ut så många som 25 % av de testade.

(2). I en urvalssituation får de som bluffat mest fortfarande en stor fördel. I den oselegerade gruppen tillhörde 6 av de 10 som hade högst värde i skönmålning de 10 främsta i testet (jagstyrka). I den selegerade gruppen var siffran densamma. Detta innebär att man fortsätter att gynna dem som bluffat i sina svar.

(3). Den multipla korrelationen mellan den overta skalan på skönmålning och hela uppsättningen personlighetsvariabler var, i hela datamängden, 0.637. Detta motsvarar 41 % gemensam varians mellan personlighetsvariablerna och måttet på skönmålning. Sedan de 25 % som hade högsta värdet i overt skönmålning uteslutits sjönk korrelationen till 0.548, gemensam varians 30 %. Detta innebär att  $\frac{3}{4}$  av effekten av skönmålning fanns kvar även om 25 % av de testade hade uteslutits på grundval av höga värden i skönmålningsskalan. Tar man hänsyn även till den variation i skönmålning som mäts med den koverta skalan kvarstår 91 % av överlappningen mellan skönmålning och personlighetsvariablerna efter det att de 25 % högsta i overt skönmålning uteslutits

Slutsatsen är att användning av ett gränsvärde och uteslutande av testpersoner med högt värde i skönmålning har en mycket begränsad effekt. Skönmålning kvarstår som problem även efter en sådan procedur. Forskningen på området har dessutom visat att strategin att utesluta vissa med höga värden i skönmålning fungerar dåligt, se Reeder och Ryan (Reeder & Ryan, 2012).

## **Varningar**

Det har gjorts ganska många försök att minska skönmålningen med hjälp av varningar till de testade, antingen om att skönmålning kan upptäckas eller om negativa konsekvenser för dem som skönmålar. En sammanställning av den forskningen tyder på att varningar ibland minskar omfattningen av skönmålning men att effekten är liten (Dwight & Donovan, 2003; Pace & Borman, 2006). Det verkar heller inte som om varningar för skönmålning ökar validiteten hos ett test (Robson, Jones, & Abraham, 2008).

## **Lura testet?**

Som vi har sett har *UPP* ett inbyggt skydd mot skönmålning som baseras på två skalor som mäter tendensen att ge taktiska svar. En intressant fråga är hur mycket testresultaten kan snedvridas av någon som har en stark inriktning mot att "lura" testet<sup>11</sup>. En sådan attityd går utöver önskan att framstå i så bra dager som möjligt, eftersom den taktiken rimligen begränsas av att testresultatet också måste ha en viss trovärdighet.

---

<sup>11</sup> I litteraturen talar man om "sabotage" av testresultat.



Jag erbjöd på Internet till en tävling mot testet. Så här såg inbjudningen ut:

### Stresstestning av UPP-testet: kan du lura testet?

Personlighetstest brukar vara lätta att lura. Vill man framstå som mycket bättre än man egentligen är så kan det ordnas lätt, eftersom få test korrigerar för skönmålning och dessutom är pinsamt lätta att genomskåda. Det gäller emellertid *inte UPP*. Vi har många skäl att tro att korrigeringen fungerar bra i *UPP*-testet men vill nu se om den fungerar också när den testade personen går in för att lura oss så mycket som möjligt. Detta blir ett stresstest av *UPP*!

Tror du att du kan lura *UPP*? Vi utmanar dig! Skicka ett mail till [henrik.nilheim@gmail.com](mailto:henrik.nilheim@gmail.com) så får du inloggningssuppgifter för att ta testet. Du får sedan en fullständig testrapport varav bl a framgår dina okorrigerade och korrigerade värden. Till den som lyckas få samma värden i dessa två fall skickar vi en skön morgonrock och två kaffemuggar, och ett diplom varav framgår att du är smartare än *UPP*-testet!

Det är bråttom! Erbjudandet gäller bara fram till 1 september. Lycka till!

21 personer ställde upp och försökte lura testet, 9 av dem svarade dessutom ärligt. Den senare gruppen är för liten för en meningsfull analys, och även de som aktivt försökte lura oss var lite färre än vi hoppats.

Fördelning av värdena i overt och kovert skönmålning framgår av Tabell 5.

Stanine	Overt	Kovert
1	6.7	6.7
2	6.7	0.0
3	6.7	0.0
4	6.7	20.0
5	13.3	16.7
6	6.7	10.0
7	10.0	10.0
8	26.7	26.7
9	16.7	10.0

I normgruppen ligger 11 % på värdena 8 och 9, här alltså 3-4 gånger fler.

Ett medelvärde av alla skalvärden beräknades, före och efter korrektion för skönmålning. Före korrektion var medelvärdet 6.2, efter 5.7<sup>12</sup>. Det var alltså ca 1/2 skalenhets skillnad.

Det fanns några testpersoner som hade låga värden i skönmålning. De fick emellertid betala ett pris för detta. De fick låga värden också i övriga skalor. För dem som hade låga värden i

<sup>12</sup> Perfektionism ej inräknad eftersom den skalan är negativt korrelerad med skönmålning,

overt skönmålning var medelvärdet av alla skalor före korrektion 4.9, för dem med låga värden i kovert skönmålning var motsvarande värde 4.5.

Den starkaste trenden i dessa data var tydlig: stark skönmålning och sänkta skalvärden därefter. Det gick inte för dessa testpersoner att lura testet. En mindre grupp tycktes försöka sig på låg skönmålning i stället, men de fick också låga värden i andra skalor. Ingen lyckades få höga testpoäng i kombination med låg skönmålning, eller behålla höga skalpoäng tack vare genomsnittlig skönmålning. *Mycket* höga testvärden var vanliga före korrektion. Det är samma observation som praktiker ofta möter vid testningar i skarpt läge, med test som är oskyddade mot skönmålning.

Gick det att lura *UPP*-testet? Svar nej, men forskning av detta slag borde helst baseras på större grupper och mera information borde insamlas om vilka strategier de testade använde sig av. Det fanns en tendens i data att en mindre andel av testpersonerna tycktes genomskåda skalan för overt skönmålning. Det är inget som kommit fram i tidigare dataanalyser, men det ger onekligen stöd för användning av en kovert skala. Den gick inte att genomskåda.

Skepsis mot personlighetstest av självrapporttyp tycks vara ganska vanligt förekommande. Det är inte svårt att inse att man kan lura dessa test och framställa sig som mycket mera personlighetsmässigt kvalificerad än man är. Saken blir inte bättre av att det bara är ett fåtal test av denna typ som används på den svenska marknaden, och att många har tagit dessa test ett stort antal gånger, och då lärt sig att utveckla strategier för att svara så att deras chanser maximeras (Hausknecht, 2010). I *UPP*-testet ställs frågan om testpersonen tagit testet tidigare. I en grupp bestående av 768 testade sökande till chefs- eller specialistjobb var fördelningen av svar, i procent:

Nej: 93.4  
En gång: 5.1  
Två gånger: 1.2  
3-5 gånger: 0.1  
>5 gånger: 0.3

Det var alltså ytterst få som tagit testet tidigare. Vad gäller andra test på marknaden tycks denna information saknas.

## Ordningen mellan testuppgifterna

Det är troligen vanligt att man ger uppgifterna i ett personlighetstest i mer eller mindre slumpmässig ordning, men det förekommer också att de ges i block bestående av alla uppgifter som ingår i en viss skala, se t ex CTI-testet (Carlstedt & Widén, 1998). Även om forskningen om denna fråga inte är omfattande tycks det som om en slumpmässig ordning motverkar skönmålning (Vasilopoulos & Cucina, 2006), kanske för att det är svårare att genomskåda testets syften med en sådan ordning.

## Effekter av skönmålning

*"It seems hasty to conclude that faking does not matter when its very presence has the potential to call into question many of the inferences that are made during the selection process"* (Peterson & Griffith, 2006) p 258

Rosse et al. (Rosse, Stecher, Miller, & Levin, 1998) visade att skönmålning förekom bland sökande, och att korrektion gav en förändrad rangordning vilket i sin tur kunde förväntas påverka beslut om anställning, mera ju mindre urvalskvoten var. En urvalskvot på minst 50 % har sagts vara nödvändig för att motverka de destruktiva effekterna av skönmålning (Griffith, Chmielowski, & Yoshita, 2007), men hur ofta arbetar man med sådana urvalskvoter? Ofta åberopade resultat av Ones et al. (Ones, et al., 1996), som tycks visa att korrektion inte påverkade testens validitet, menar de är missvisande eftersom den metodik de använde (korrelationer mellan test och kriterier) var okänslig. Den viktiga gruppen är den mindre del som har de högsta testpoängen och om man inte korrigerar kommer den att bestå till stor del av testade som har skönmålat starkt<sup>13</sup>.

Ett i debatten förekommande påstående är att den som skönmålar "vet vad som krävas i jobbet", och därför skulle också prestera bättre i jobbet. Mycket långsökt! Se t ex detta citat:

*"It is far from clear that someone motivated to make a good impression on a personality test when applying for a job will be as highly motivated to make a good impression after being hired, especially over intervals on which most performance criteria are based (i.e., 6 to 12 months)".* (Tett, et al., 2006)

Betydande ekonomiska vinster kan göras med korrektion för skönmålning, även om effekterna på validiteten är små (Burns & Christiansen, 2006; Choi, Kluemper, & Sauley, 2011). Om inget görs åt skönmålningen riskerar man att de som får bäst testresultat till stor del består av dem som bluffat starkt och ohämmat på testet, och att testets validitet minskats (Holden, 2007; Topping & O'Gorman, 1997). Många har varnat för användning av test som är oskyddade mot skönmålning (McFarland, 2003). Lukoff (Lukoff, 2012) påpekar:

*"Rewarding faking with a job or a promotion sends the message that such behavior is just not condoned but encouraged, and that it is acceptable to do whatever it takes to win, even if "what it takes" involves misrepresentation or outright lying. Putting even a small number of people that believe this into a position of power – cf. Ken Lay or Bernie Madoff – can be (and has been) disastrous".* (p.250)

---

<sup>13</sup> Mina resonemang förutsätter att den strategi som används är "select in", dvs. man letar efter de bäst kvalificerade i en grupp av sökande. Då är skönmålning av särskilt stor betydelse. Om testet i stället används för "select out", dvs. man letar efter vilka som ska uteslutas från fortsatt prövning, är skönmålning troligen av mindre betydelse eftersom denna grupp, om den är tillräckligt liten, troligen inte innehåller många som bluffat på testet. Men i så fall finns problemet med skönmålning kvar i gruppen som går vidare.

## Skönmålning och femfaktormodellen

Femfaktormodellen (B5) har kommit att dominera personlighetstestning inom arbetslivet. Skälet är knappast att den har hög validitet. Barrick, Mount och Judge publicerade år 2001 en översikt av de meta-analyser av sambandet mellan B5-dimensioner och arbetsresultat som då fanns tillgängliga. (Senare forskning har gett likartade resultat). För övergripande analys av arbetsresultat mot B5-dimensioner rapporterade de följande genomsnittliga validiteter, Tabell 6.

Tabell 6. Genomsnittliga korrelationer mellan personlighetsdimensioner och arbetsresultat Baseras på 5 meta-analyser (4 för öppenhet).

Dimension	Korrelation
Emotionell stabilitet	0.12
Extraversion/Utåtvändhet	0.12
Öppenhet	0.05
Vänlighet/Följsamhet	0.10
Noggrannhet	0.23

Detta kan tjäna som en sammanfattning av alla meta-analyser av B5-dimensioner mot arbetsresultat. Vi ser att denna typ av analys ger en tydlig validitet enbart för noggrannhet och inte ens denna är särskilt stor. Det är 20 års forskning som ligger bakom siffrorna och hundratals studier. Det kan tilläggas att B5-testen (NEO-PI-R, HPI, OPQ, MINT m fl) nästan alltid är oskyddade mot skönmålning, eller endast har ineffektivt skydd. Det innebär att många beslut som fattas på grundval av sådana test troligen inte är optimala.

## Normering

Alla test måste ha normer, bestående av testresultat i en eller flera relevanta jämförelsegrupper. Om man inte tar hänsyn till skönmålning kan jämförelser bli ytterst missvisande, om nämligen normdata är insamlade i oskarpt läge och aktuella testningar i skarpt. Detta är troligen ett mycket vanligt fall (Dilchert & Ones, 2012). Ett exempel är den ursprungliga svenska versionen av NEO-PI-R (Bergman, 2000), att annat är PJP (Sjöberg, Sjöberg, & Forssén, 2006). Skillnaden mellan en oskarp norm och testdata insamlade i skapt läge kan vara mycket stor Skillnaden försvinner efter korrektion för skönmålning. Detta resultat betyder att normdatabasen är relevant även vid arbete i skarpt läge, om man nämligen korrigerar på det sätt som görs i *UPP*-testet. *Om man inte korrigerar men ändå använder en oskarp norm kommer man att systematiskt och kraftigt överskatta testresultat som insamlats i skapt läge.*

## Upprepade testningar och skönmålning

En av ytterst få forskare som påstår att skönmålning inte förekommer i skarpa lägen är Hogan (Hogan, Barrett, & Hogan, 2007). Hogan et al. genomförde ett testprogram två gånger, den andra gången med sökande som hade misslyckats vid det första tillfället. De fick emellertid

inga förhöjda värden i personlighetsskalor, vilket man kunde ha väntat sig om de testade hade tillgripit skönmålning. Hogan et al. drog starka slutsatser av detta (enstaka) resultat och skriver:

*”When people complete a well-validated personality measure as part of a job application process, are denied employment, reapply some time later, and then take the personality measure a second time, their scores will not change significantly on the second occasion. This claim is fully supported by our data. We also believe it is reasonable to assume that the applicants tried to improve their scores on the second occasion; on the basis of this assumption, we interpret the data as showing that, when they try, applicants are unable to improve their scores substantially.” (p. 1282)*

Sackett (Sackett, 2012) påpekar emellertid att andra ej kunnat replikera detta resultat och att en förklaring kan vara att personerna i studien av Hogan et al. tillskrev sitt misslyckande till andra delar av testprogrammet (kognitiva test). Det har påvisats många gånger att sökande kan förbättra sina resultat på personlighetstest av självrapporttyp. Skälet kan också helt enkelt vara att de skönmålade så mycket vid första tillfället att de inte kunde höja nivån märkbart. Det är ganska troligt med tanke på hur vanlig och stark skönmålning är.

## Debatten om skönmålning

Debatten om skönmålning är intensiv. Det finns starka motsättningar mellan anhängarna av olika ståndpunkter. Man kan urskilja följande grupper av argument:

*Skönmålning är vanlig och stark, kan inte motverkas eller korrigeras, och gör testen värdelösa.* Detta är troligen en ganska vanlig uppfattning, särskilt bland personer som inte har speciellt djupa kunskaper inom området. Det är ju enkelt att inse att testpersoner kan svara oärligt, och att inget hindrar en del av dem från att göra det om mycket står på spel (skarpt läge). Mot detta kan man anföra att det finns forskning på området som ger stöd för en mera nyanserad uppfattning och som visat att det faktiskt går att komma tillrätta med problemet – kanske inte helt, men i stor utsträckning.

*Skönmålning förekommer inte eller bara ytterst sällan, alltså är hela frågan utan betydelse.* Detta är en uppfattning som bygger på enstaka, avvikande forskningsresultat som tycks ha visat att t ex sökande eller anställda besvarat ett test på mycket likartade sätt. Eftersom omfattande forskning visat att de i normalfallet inte alls svarar på ett likartat sätt är ståndpunkten ohållbar.

*Skönmålning förekommer ibland men saknar betydelse för det praktiska värdet av testen.* Detta påstående bygger på viss forskning som tycks ha visat att testen inte har lägre validitet om skönmålning förekommer, eller att korrektion för skönmålning inte höjer validiteten. Emellertid bortser man då från att mycket stora effekter ofta uppkommer till följd av skönmålning i individuella fall, att det handlar om en "tätgrupp" i testet och att den övergripande korrelationen mellan ett test och ett kriterium på arbetsresultat är ett mycket okänsligt mått på effekter av skönmålning.

*Skönmålning förekommer men slutsatser av testen bör grundas på "profiler" och sådana kan inte förvrängas med hjälp av skönmålning.* "Profiler" kan mycket väl påverkas av skönmålning. Effekterna av skönmålning är ju inte generella utan varierar över skalor.

*Skönmålning kan effektivt motverkas med hjälp av ipsativa svarsformat. Ipsativa format motverkar inte skönmålning effektivt och har ett antal nackdelar som beskrivits i texten ovan.*

*Skönmålning kan effektivt motverkas med hjälp av varningar. Varningar har till en del de avsedda effekterna att minska skönmålning men effekterna är svaga och det mesta av problemet kvarstår.*

*Skalor som mäter "social önskvärdhet" är i själva verket mätningar av "personlighet". Detta argument är ohållbart av två skäl. För det första är den personlighetsdimension som vanligen avses "need for approval", men oberoende mått på denna dimension har inget samband med den traditionella skalan för att mäta social önskvärdhet (Marlowe-Crowne). För det andra saknas logik i argumentet att en skala invalideras om den korrelerar med något annat som inte var avsett eller förutsett. Antag att en skala för mätning skönmålning råkar samvariera med kroppsvikt eller, för den delen, intelligens. Är det ett bevis för att skalan har misslyckats i sitt syfte? Det är helt enkelt irrelevant för frågan om skalan mäter det som den är avsedd att mäta. Validering av skalan kan ske med empirisk forskning av den typ som redovisats i denna rapport.*

*Skönmålning bör mätas som skillnaden mellan två testtillfällen, ett med instruktionen att skönmåla ett med normal instruktion att svara ärligt. Detta mått har visat sig ha svårgenomskådade egenskaper. Att skönmåla när man är direkt uppmanad till att göra det förefaller vara något annat än att göra det spontant i en skarp situation i syfte att förbättra sitt testresultat. I praktiken kan inte denna metodik användas för att komma till rätta med skönmålning i en skarp situation, t ex testning vid jobbansökan.*

*Om skönmålning förekommer kan vi inte veta eftersom vi inte känner de "sanna" värdena i personlighet; inte heller kan vi motverka eller korrigera för skönmålning. Detta är ett uttryck för en kunskapsteoretisk pessimism som kanske kan försvaras på filosofiska grunder men som inte tycks leda till några konstruktiva insikter eller till problemlösning.*

## **Diskussion och slutsatser**

Den metodik för att motverka effekterna av skönmålning som förfäktas i denna rapport bygger på idéer som funnits länge, men den är vidareutvecklad här.

1. Korrigeringen för skönmålning görs separat för varje testskala med modeller som baseras på empiriska data. Olika skalor kräver olika grad av korrektion.
2. Skönmålning mäts inte bara på traditionellt sätt med en skala för "social önskvärdhet" utan också med en kovert skala som inte kan genomskådas av testpersonerna. De två skalorna har gett mycket likartade resultat.

Omfattande validering av detta angreppssätt har redovisats i denna rapport:

1. Korrigeringen eliminerar ca 95 % av effekterna av konsekvenserna av testningen, som i sin tur påvisas ha mycket stor effekt på graden av skönmålning. Skarpt läge ger mycket mera skönmålning än oskarpt läge; olika grader av skärpa hos läget har olika starka effekter.

2. Efter skönmålning ökar ibland validiteten hos testskalorna, ibland minskar den något eller är oförändrad.
3. Tiden för att besvara en uppgift i testet ökar hos dem som skönmålar mera. Detta är i överensstämmelse med antagandet att skönmålning kräver mera eftertanke än ett spontant svar.
4. Om testpersonerna direkt tillfrågas om de har skönmålat ser vi högre grad av skönmålning hos dem som svarar ja på frågan, *men också hos dem som absolut förnekar att de skönmålat.*

Mer eller mindre traditionella argument mot vårt angreppssätt diskuteras i rapporten i det närmast föregående avsnittet och befinns vara ohållbara. Ståndpunkten i denna rapport är att skönmålning ofta förekommer, är stark och kan effektivt elimineras med hjälp av korrigering baserad på särskilda mätningar av ”social önskvärdhet”.

Testbranschen måste ta fram trovärdiga och validerade lösningar av problemet med skönmålning. I själva verket tycks många testleverantörer använda någon form av skalor som mäter skönmålning, men i de flesta fallen tycks ingen effektiv metod användas för att utnyttja den information de ger (Goffin & Christiansen, 2003). Om inget görs åt skönmålning kommer testningar att i ökande grad upplevas som godtyckliga och ett sätt för bluffande personer att ta sig fram. Detta gäller även om man kan hänvisa till forskning som tyder på att validiteten hos test i stort inte påverkas av skönmålning. Även om detta är sant (det kan diskuteras) gäller fortfarande att oskyddade test prioriterar dem som bluffar i sina svar. Testleverantörer måste kunna visa att de har evidens för sina test, som framtagits i kompetent genomförd forskning. Denna evidens måste hållas tillgänglig för alla som vill ta ställning till testens kvalitet, till rimligt pris och helst vara gratis och öppen för alla<sup>14</sup>.

Den av Ziegler, Maccann och Roberts redigerade boken (Ziegler, et al., 2012) som nämndes i inledningen till denna rapport innehåller ett antal intressanta förslag till nya metoder för att upptäcka om testpersoner skönmålar, men är alla ännu på ett för tidigt stadium för att vara praktiskt användbara; inte heller har det påvisats att de fungerar bättre än mer eller mindre traditionella mått på skönmålning. Några exempel:

1. Paulhus skriver om ”overclaiming” (Paulhus, 2012), en term som syftar på att man frågar en person om han eller hon känner till ett antal uppräknade begrepp, t ex författare. Vissa av begreppen i listan är ”kuggfrågor”, de existerar inte i verkligheten. Om man ändå påstår sig känna till dem kan det vara ett tecken på att man bluffar. Det finns ett visst stöd för denna teknik men den är inte praktiskt användbar i testsammanhang.
2. Ett kapitel av Zickar och Sliter ger en översikt av de försök som gjorts att mäta skönmålning genom att formulera och testa modeller för testbeteende (”item response models”) (Zickar & Sliter, 2012). Tanken bakom denna ansats är att testpersoner vars data inte passar väl in på modellerna kan vara sådana som bluffar. Det visar sig att framgången med denna ansats är mycket begränsad; den är inte praktiskt användbar och inte bättre än traditionella metoder.

---

<sup>14</sup> . En rundringning till svenska testleverantörer våren 2013 visade att inga var villiga att lämna ut manualer annat än till kunder som köpt licenser. Samma tycks gälla forskningsrapporter, i den mån sådana finns.

3. Användning av reaktionstidsmått på liknande sätt som i attitydmätning kan kanske vara ett sätt att eliminera eller minska effekterna av skönmålning, men är inte en praktiskt användbar metodik.
4. Bedömningar av personlighet av andra personer som känner den som bedöms kan eventuellt vara mindre behäftade med skönmålning, men är knappast en praktiskt användbar lösning i de flesta fall.

## **Uppsummering**

Vilka slutsatser leder denna rapport till, speciellt med tanke på den vanliga kritik av år metodik som nyligen upprepats i boken redigerad av Ziegler et al. (Ziegler, et al., 2012)? Låt mig ta diskussionen steg för steg.

6. Förekommer skönmålning, är den vanlig och stark? Svaret är ett entydigt ja.
7. Har skönmålning effekter på beslutsfattandet på grundval av testning, t ex vid urval? Svaret är ja.
8. Här skönmålning betydelse för normdata? Svaret är ja, om testet ska användas i skarpt läge måste normdata vara insamlade i skarpt läge.
9. Kan skönmålning undvikas, åtminstone till betydande del, med hjälp av
  - a. Varningar
  - b. Omformulering av item till att bli mera neutrala
  - c. Särskilda mätningar av tendensen att tillskriva sig mera kunskaper än man har
  - d. Anpassning av matematiska modeller till testdata och mätning av i vilken mån testpersoner avviker från dessa modeller

Svaret för alla dessa ansatser är nej, de löser inte problemet

10. Ett ganska vanligt sätt att hantera skönmålning tycks vara att utesluta de testpersoner som har högst värde i en sådan skala, t ex de 25 % som ligger högst. Detta har en svag och otillräcklig effekt, se Reeder och Ryan (Reeder & Ryan, 2012).

Ännu tycks det inte finnas någon ny metodik för att upptäcka och korrigera för skönmålning som är tillräckligt lovande för att leva upp till kraven att den ska vara teoretiskt och empiriskt väl underbyggd, och praktiskt användbar. Korrektur med en särskild skala som mäter skönmålning fungerar, även om inte heller denna metod helt eliminerar effekterna.



## Referenser

- Arvonen, J. (2002). *Change, production and employees. An integrated model of leadership*. Stockholm: Department of Psychology, University of Stockholm.
- Barger, V. A. (2012). *Individual differences in need for approval: Measurement and marketing implications*. ProQuest Information & Learning, US.
- Bergman, H. (2000). *Manual NEO PI-R*. Stockholm: Hogrefe.
- Bernard, M. E. (1998). Validation of the General Attitude and Belief Scale. *Journal of Rational-Emotive and Cognitive-Behavior Therapy*, 16(3), 183-196.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A Meta-Analytic Investigation of Job Applicant Faking on Personality Measures. [doi:10.1111/j.1468-2389.2006.00354.x]. *International Journal of Selection and Assessment*, 14(4), 317-335.
- Boon, J., Gozna, L., & Hall, S. (2008). Detecting 'faking bad' on the Gudjonsson Suggestibility Scales. [doi:10.1016/j.paid.2007.08.005]. *Personality and Individual Differences*, 44(1), 263-272.
- Bowen, C.-C., Martin, B. A., & Hunt, S. T. (2002). A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires. *International Journal of Organizational Analysis*, 10(3), 250-259.
- Bradley, K. M., & Hauenstein, N. M. A. (2006). The moderating effects of sample type as evidence of the effects of faking on personality scale correlations and factor structure. *Psychology Science*, 48(3), 313-335.
- Brown, A., & Bartram, D. (2009). *Development and psychometric properties of the OPQ32r. Supplement to the OPQ 32 technical manual: SHL*.
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71(3), 460-502.
- Burns, G. N., & Christiansen, N. D. (2006). Sensitive or senseless. On the use of social desirability measures in selection and assessment. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 115-150). Greenwich, CT: IAP.
- Carlstedt, L., & Widén, H. (1998). *CTI – Commander Trait Inventory. Manual*. Karlstad: LI: Försvarshögskolan.
- Choi, S., Kluemper, D. H., & Sauley, K. S. (2011). What if we fake emotional intelligence? A test of criterion validity attenuation. [doi:10.1080/00223891.2011.558870]. *Journal of Personality Assessment; Journal of Personality Assessment*, 93(3), 270-277.
- Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering Forced-Choice Item Formats for Applicant Personality Assessment. *Human Performance*, 18, 267-307.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2006). Forcing choices in personality measurement: Benefits and limitations. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 263-282). Greenwich, CT: IAP.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting and Clinical Psychology*, 24, 349-354.
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York: Wiley.

- De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. [doi:10.1037/a0014211]. *Psychological Bulletin*, *135*(3), 347-368.
- Dilchert, S., & Ones, D. S. (2012). Application of preventive strategies. In M. Ziegler, C. Maccann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 177-200). New York: Oxford University Press.
- Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An Assessment of the Prevalence, Severity, and Verifiability of Entry-Level Applicant Faking Using the Randomized Response Technique. *Human Performance*, *16*(1), 81-106.
- Dwight, S. A., & Donovan, J. J. (2003). Do Warnings Not to Fake Reduce Faking? *Human Performance*, *16*(1), 1-23.
- Eitan, D. G. (2011). Is the truth visible? Study along the L scale. *The Open Psychology Journal*, *4*.
- Ellingson, J. E. (2012). People fake only when they need to fake. In M. Ziegler, C. Maccann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 19-33). New York: Oxford University Press.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, *84*(2), 155-166.
- 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.
- Fernando, P. J., & Chico, E. (2001). Detecting dissimulation in personality test scores: A comparison between person-fit indices and detection scales. *Educational and Psychological Measurement*, *61*, 997-1012.
- Ferrando, P. J., & Anguiano-Carrasco, C. (2011). Faking propensity and faking-related change: A model-based analysis of the EPQ-R scores. [doi:10.1016/j.paid.2011.05.006]. *Personality and Individual Differences*, *51*(4), 497-501.
- Gibby, R. E., & Zickar, M. J. (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. [doi:10.1037/a0013041]. *History of Psychology*, *11*(3), 164-184.
- Goffin, R. D., & Christiansen, N. D. (2003). Correcting Personality Tests for Faking: A Review of Popular Personality Tests and an Initial Survey of Researchers. [doi:10.1111/j.0965-075X.2003.00256.x]. *International Journal of Selection and Assessment*, *11*(4), 340-344.
- Griffith, R., Malm, T., English, A., Yoshita, Y., & Gujar, A. (2006). Applicant faking behavior. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 151-178). Greenwich, CT: IAP.
- Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, *36*(3), 341-355.
- Griffith, R. L., & Converse, P. D. (2012). The rules of evidence and the prevalence of applicant faking. In M. Ziegler, C. Maccann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment*. Oxford: Oxford University Press.
- Griffith, R. L., & Peterson, M. H. (Eds.). (2006). *A closer examination of applicant faking behavior*. Greenwich, CT: IAP.
- Harland, L. K. (2003). Using Personality Tests in Leadership Development: Test Format Effects and the Mitigating Impact of Explanations and Feedback. *Human Resource Development Quarterly*, *14*(3), 285-301.

- Hausknecht, J. P. (2010). Candidate persistence and personality test practice effects: Implications for staffing system management. [doi:10.1111/j.1744-6570.2010.01171.x]. *Personnel Psychology*, 63(2), 299-324.
- Heggestad, E. D. (2012). A conceptual representation of faking. . In M. Ziegler, C. Maccann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 87-101). New York: Oxford University Press.
- Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. [doi:10.1037/0021-9010.91.1.9]. *Journal of Applied Psychology*, 91(1), 9-24.
- Hendrickson, G. (1934). Some assumptions involved in personality measurement. *Journal of Experimental Education*, 2, 243-249.
- Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 74, 167-184.
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92, 1270-1285.
- Holden, R. R. (1998). Detecting fakers on a personnel test: Response latencies versus a standard validity scale. *Journal of Social Behavior & Personality*, 13(2), 387-398.
- Holden, R. R. (2007). Socially desirable responding does moderate personality scale validity both in experimental and in nonexperimental contexts. [doi:10.1037/cjbs2007015]. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 39(3), 184-201.
- Holden, R. R., & Hibbs, N. (1995). Incremental validity of response latencies for detecting fakers on a personality test. [doi:10.1006/jrpe.1995.1021]. *Journal of Research in Personality*, 29(3), 362-372.
- Holden, R. R., & Kroner, D. G. (1992). Relative efficacy of differential response latencies for detecting faking on a self-report measure of psychopathology. [doi:10.1037/1040-3590.4.2.170]. *Psychological Assessment*, 4(2), 170-173.
- Hough, L. M. (1998). Personality at work: Issues and evidence. In M. D. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 131-166). Mahwah, NJ: Erlbaum.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75, 581-595.
- Hurtz, G. M., & Alliger, G. M. (2002). Influence of coaching on integrity test performance and unlikely virtues scale scores. [doi:10.1207/S15327043HUP1503\_02]. *Human Performance*, 15(3), 255-273.
- Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? [doi:10.1207/S15327043HUP1304\_3]. *Human Performance*, 13(4), 371-388.
- Komar, S., Komar, J. A., Robie, C., & Taggar, S. (2010). Speeding personality measures to reduce faking: A self-regulatory model. [doi:10.1027/1866-5888/a000016]. *Journal of Personnel Psychology*, 9(3), 126-137.
- Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: Development and validation of an interview faking behavior scale. [doi:10.1037/0021-9010.92.6.1638]. *Journal of Applied Psychology*, 92(6), 1638-1656.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1, 27-66.

- Lindner, H., Kirkby, R., Wertheim, E., & Birch, P. (1999). A brief assessment of irrational thinking: The shortened General Attitude and Belief Scale. [doi:10.1023/A:1018741009293]. *Cognitive Therapy and Research*, 23(6), 651-663.
- Lukoff, B. (2012). Is faking inevitable? Person-level strategies for reducing faking. In M. Ziegler, C. Maccann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 240-252). New York: Oxford University Press.
- 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.
- Mabon, H. (2004). *Arbetspsykologisk testning. Om urvalsmetoder i arbetslivet. Andra upplagan*. Stockholm: Psykologiförlaget.
- Magnusson, D. (2003). *Testteori*. Stockholm: Psykologiförlaget.
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. [doi:10.1037/0022-006X.51.6.882]. *Journal of Consulting and Clinical Psychology*, 51(6), 882-888.
- McFarland, L. A. (2003). Warning Against Faking on a Personality Test: Effects on Applicant Reactions and Personality Test Scores. [doi:10.1111/j.0965-075X.2003.00250.x]. *International Journal of Selection and Assessment*, 11(4), 265-276.
- 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.
- Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. [doi:10.1037/h0053634]. *Journal of Applied Psychology*, 30(5), 525-564.
- Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, 11(2/3), 245-269.
- 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.
- Pace, V. L., & Borman, W. C. (2006). The use of warnings to discourage faking on noncognitive inventories. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 283-304). Greenwich, CT: IAP.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social-psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.
- Paulhus, D. L. (2012). Overclaiming on personality questionnaires. In M. Ziegler, C. Maccann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 151-164). New York: Oxford University Press.
- Paulhus, D. L., & Reid, D. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, 60, 307-317.
- Pauls, C. A., & Crost, N. W. (2005). Effects of different instructional sets on the construct validity of the NEO-PI-R. [doi:10.1016/j.paid.2005.01.003]. *Personality and Individual Differences*, 39(2), 297-308.
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research. Explanation and prediction*. New York: Holt, Rinehart and Winston.
- Peterson, M. H., & Griffith, R. L. (2006). Faking and job performance. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 233-261). Greenwich, CT: IAP.

- Reeder, M. C., & Ryan, A. M. (2012). Methods for correcting for faking. In M. Ziegler, C. Maccann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 131-150). New York: Oxford University Press.
- Rees, C. J., & Metcalfe, B. (2003). The faking of personality questionnaire results: Who's kidding whom? *Journal of Managerial Psychology*, *18*(2), 156-165.
- Robson, S. M., Jones, A., & Abraham, J. (2008). Personality, faking, and convergent validity: A warning concerning warning statements. [doi:10.1080/08959280701522155]. *Human Performance*, *21*(1), 89-106.
- Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement*, *20*(4), 335-343.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. [doi:10.1037/0021-9010.83.4.634]. *Journal of Applied Psychology*, *83*(4), 634-644.
- Röhner, J., Schröder-Abé, M., & Schütz, A. (2011). Exaggeration is harder than understatement, but practice makes perfect! Faking success in the IAT. [doi:10.1027/1618-3169/a000114]. *Experimental Psychology*, *58*(6), 464-472.
- Sackett, P. R. (2012). Faking in personality assessments. Where do we stand? In M. Ziegler, C. Maccann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 330-344). New York: Oxford University Press.
- Sackett, P. R., & Lievens, F. (2008). Personnel selection. [doi:10.1146/annurev.psych.59.103006.093716]. *Annual Review of Psychology*, *59*, 419-450.
- Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. [doi:10.1037/0021-9010.78.6.966]. *Journal of Applied Psychology*, *78*(6), 966-974.
- Sjöberg, A., Sjöberg, S., & Forssén, K. (2006). *Predicting Job Performance, PJP Manual*. Stockholm: Assessio.
- Sjöberg, L. (2005). Självskattningstest: Den som ljuger fräckast vinner. *Personal och Ledarskap*, *36*(6), 22-25.
- Sjöberg, L. (2008). *Bortom Big Five: Konstruktion och validering av ett personlighetstest*. SSE/EFI Working Paper Series in Business Administration No. 2008:7). Stockholm: Stockholm School of Economics.
- Sjöberg, L. (2009). *UPP-testet: Korrektion för skönmålning. (The UPP test: Correction for impression management)*. *Forskningsrapport 2009:3*. Stockholm: Psykologisk Metod AB.
- Sjöberg, L. (2010). *UPP-testet: Mångfald gynnas av korrektion för skönmålning. Forskningsrapport 2010: 2*. Stockholm: Psykologisk Metod AB.
- Sjöberg, L. (2011). *UPP-testet. Teknisk manual*. Stockholm: Psykologisk Metod AB.
- Sjöberg, L., Bergman, D., Lornudd, C., & Sandahl, C. (2011). *Sambandet mellan ett personlighetstest och 360-graders bedömningar av chefer i hälso- och sjukvården*. Stockholm: Karolinska Institutet, Institutionen för lärande, informatik, management och etik (LIME).
- Sjöberg, L., & Wolgers, G. (2012). *Personlighetstestning vid antagning till officersutbildning. En jämförelse av CTI och UPP-testet i olika grader av "skarpt läge". Rapport F: 39*. Karlstad: Försvarshögskolan, Institutionen för säkerhet, strategi och ledarskap.
- Tett, R. P., Anderson, M. G., Ho, C.-L., Yang, T. S., Huang, L., & Hanvongse, A. (2006). Seven nested questions about faking on personality tests. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 43-83). Greenwich, CT: IAP.

- Tonković, M. (2012). Are there personality traits that predispose applicants to fake noncognitive measures in personnel selection? *Review of Psychology, 19*(1), 29-36.
- Topping, G. D., & O'Gorman, J. G. (1997). Effects of faking set on validity of the NEO-FFI. [doi:10.1016/S0191-8869(97)00006-8]. *Personality and Individual Differences, 23*(1), 117-124.
- Uziel, L. (2010). Rethinking social desirability scales: from impression management to interpersonally oriented self-control. *Perspectives on Psychological Science, 5*(3), 243-262.
- van Hooft, E. A. J., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. [doi:10.1037/a0025711]. *Journal of Applied Psychology, 97*(2), 301-316.
- Vasilopoulos, N. L., & Cucina, J. M. (2006). Faking on noncognitive measures. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 305-331). Greenwich, CT: IAP.
- Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-Choice Personality Tests: A Measure of Personality and Cognitive Ability? [doi:10.1207/s15327043hup1903\_1]. *Human Performance, 19*(3), 175-199.
- Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. [doi:10.1016/j.jrp.2005.03.003]. *Journal of Research in Personality, 40*(5), 472-481.
- Weed, N. C., Ben-Porath, Y. S., & Butcher, J. N. (1990). Failure of Wiener and Harmon Minnesota Multiphasic Personality Inventory (MMPI) subtle scales as personality descriptors and as validity indicators. [doi:10.1037/1040-3590.2.3.281]. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*(3), 281-285.
- Weissman, A. N., & Beck, A. (1978). *Development and validation of the dysfunctional attitudes scale: A preliminary investigation*. Paper presented at the Proceedings of the meeting of the American Educational Research Association.
- Visweswaran, C., & Ones, D. S. (1999). Meta-analysis of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197-210.
- Vrij, A. (2000). *Detecting lies and deceit*. New York: Wiley.
- Zickar, M. J., & Gibby, R. F. (2006). A history of faking and socially desirable responding on personality tests. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 21-42). Greenwich, CT: IAP.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. [doi:10.1037/0021-9010.84.4.551]. *Journal of Applied Psychology, 84*(4), 551-563.
- Zickar, M. J., & Sliter, K. A. (2012). Searching for unicorns. Item response theory-based solutions to the faking problem. In M. Ziegler, C. Maccann & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 113-130). New York: Oxford University Press.
- Ziegler, M., Maccann, C., & Roberts, R. D. (2012). *New perspectives on faking in personality assessment*. New York: Oxford University Press.

## **Appendix 1: Marlowe-Crownes skala och två skalor som mäter "Need for approval"**

### **The thirty-three items of the Marlowe-Crowne Social-Desirability scale**

1. Before voting I thoroughly investigate the qualifications of all the candidates. (T)
2. I never hesitate to go out of my way to help someone in trouble. (T)
3. It is sometimes hard for me to go on with my work if I am not encouraged. (F)
4. I have never intensely disliked anyone. (T)
5. On occasion I have had doubts about my ability to succeed in life. (F)
6. I sometimes feel resentful when I don't get my way. (F)
7. I am always careful about my manner of dress. (T)
8. My table manners at home are as good as when I eat out in a restaurant. (T)
9. If I could get into a movie without paying and be sure I was not seen I would probably do it. (F)
10. On a few occasions, I have given up doing something because I thought too little of my ability (F)
11. I like to gossip at times. (F)
12. There have been times when I felt like rebelling against people in authority even though I knew they were right. (F)
13. No matter who I'm talking to, I'm always a good listener. (T)
14. I can remember "playing sick" to get out of something. (F)
15. There have been occasions when I took advantage of someone. (F)
16. I'm always willing to admit it when I make a mistake. (T)
17. I always try to practice what I preach. (T)
18. I don't find it particularly difficult to get along with loud mouthed, obnoxious people. (T)

19. I sometimes try to get even rather than forgive and forget. (F)
20. When I don't know something I don't at all mind admitting it. (T)
21. I am always courteous, even to people who are disagreeable. (T)
22. At times I have really insisted on having things my own way. (F)
23. There have been occasions when I felt like smashing things. (F)
24. I would never think of letting someone else be punished for my wrongdoings. (T)
25. I never resent being asked to return a favor. (T)
26. I have never been irked when people expressed ideas very different from my own. (T)
27. I never make a long trip without checking the safety of my car. (T)
28. There have been times when I was quite jealous of the good fortune of others. (F)
29. I have almost never felt the urge to tell someone off. (T)
30. I am sometimes irritated by people who ask favors of me. (F)
31. I have never felt that I was punished without cause. (T)
32. I sometimes think when people have a misfortune they only got what they deserved. (F)
33. I have never deliberately said something that hurt someone's feelings. (T)

**Demand for Approval items of the General Attitude and Belief scale (GABS)**

I can't stand being disliked by people who are important to me, and it is unbearable if they dislike me.

It's awful to be disliked by people who are important to me, and it is a catastrophe if they don't like me.

I must be liked by important people, and I will not accept not being liked by them.

I can't stand being disliked by certain people, and I can't bear the possibility of their disliking me.

It's essential to be liked by important people, and I will not accept their not liking me.



I must be liked and accepted by people I want to like me, and I will not accept their not liking me.

When people who I want to like me disapprove of me or reject me, I can't bear their disliking me.

### **Items comprising the DAS need for approval dimension**

What other people think about me is very important.

If others dislike you, you cannot be happy.

I can enjoy myself even when others do not like me.

I do not need other people's approval for me to be happy.

My value as a person depends greatly on what others think of me.

I cannot be happy unless most people I know admire me.

A person doesn't need to be well liked in order to be happy.

I can find happiness without being loved by another person.

If people consider me unattractive it need not upset me.

If people whom I care about do not care for me, it is awful.

It is awful to be disapproved by people important to you.

If you don't have other people to lean on, you are bound to be sad.

My happiness depends more on other people than it does on me.

Being isolated from others is bound to lead to unhappiness.

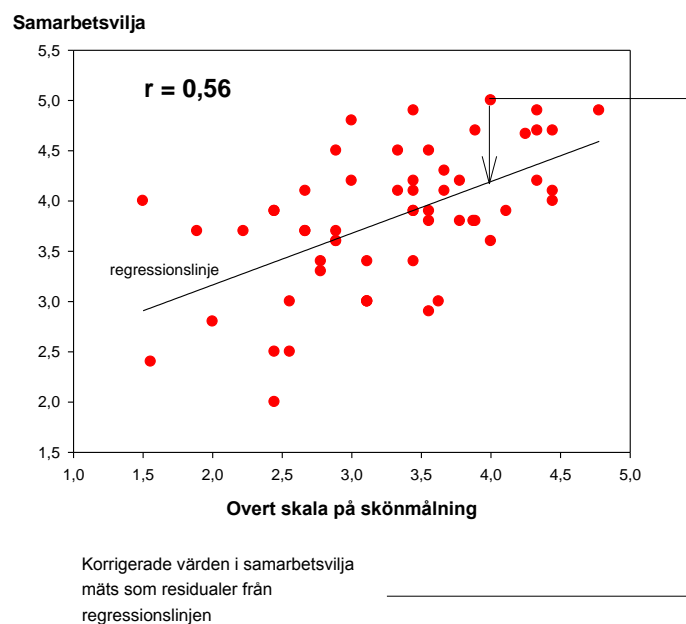
## Appendix 2. Korrektion för skönmålning: metodik

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 skönmålningskalan sällas bort och kommer inte ifråga för antagning eller anställning.

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

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



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

Beräkningen av residualer görs enligt vanlig regressionsanalys, dvs

Residual = råvärde – predicerat värde på grundval av modellen

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

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

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

där:

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

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

$e$  är en felterm.

Det från modellen predicerade testvärdet erhålls genom:

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

Residualen är:

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

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

### Estimering av regressionsvikter

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

$$\beta_1 = (r_{y1} - r_{y2} r_{12}) / (1 - r_{12}^2)$$

$$\beta_2 = (r_{y2} - r_{y1} r_{12}) / (1 - r_{12}^2)$$

Där:

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

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

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

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