

# Tester, jämställdhet och mångfald

AOnet, Umeå universitet  
26-27 september 2011

Professor Lennart Sjöberg  
Handelshögskolan i Stockholm  
och  
Psykologisk Metod AB

# Inriktning på föredraget

- Jag börjar med ett allmänt avsnitt om vissa viktiga principer vid användning och bedömning av test
- Teoretisk och empirisk analys presenteras sedan, som ger stöd för att det går att lösa svåra praktiska problem vid psykologisk testning
- Exempelen hämtas från min egen forskning med *UPP*-testet

# Forskning som underlag för denna presentation

- Data som insamlats med *UPP*-testet (mera om det senare) i fullständig version och i kortfattad screeningversion: totalt ca 8500 personer testade varav ca 80 % i "skarpt läge" (jobsökande)
- Varav 349 chefskandidater, 6116 tog screeningtestet
- Vår senaste forskningsstudie omfattade 110 deltagare med varierande bakgrund, brett upplagd valideringsstudie, bl a mot "livsframgång"

# Några principer vid testanvändning och bedömning av test

- Linjära samband med kriterier (det man vill göra en prognos av, t ex arbetsresultat)
- Indexbildning för kraftfulla prognoser
- Urvalskvoten – ju mindre desto bättre
- Evidensbaserade test
- STP:s (och andras) testgranskningar i ljuset av dessa principer

# Linjära samband

- Mycket omfattande forskning visar att samband mellan testskalor och kriterier mycket väl kan beskrivas som linjära:
  - Ju högre emotionell stabilitet, desto bättre prognos i de allra flesta jobb
  - Ju högre perfektionism, desto sämre prognos i de allra flesta jobb
- Den mest aktuella forskningen som stöder linjära samband har utförts för OPQ-testet
- Det är ett kraftigt stöd för att tänka på testdimensioner i termer av "bättre" eller "sämre" prognos - inte att "det är inte bra eller dåligt om man har sämre eller högre värde"

# Liten utveckling om sambandsstorlekar

- Korrelationer mellan testskalor och kriterier är ett naturligt och bra sätt att värdera testvaliditet
- Dessa är tyvärr ganska ofta låga
- Om man i stället gör en "signifikanstestning" av sambandet kan man få låga p-värden om man har stora stickprov
- Ett sätt att marknadsföra sig – givet att den potentielle användaren inte genomskådar detta enkla knep
- Förekommer det verkligen? Ja, just det!

# Index

- Sambanden med kriterier är vanligen av måttlig storlek för enstaka testskalor (0.1 – 0.3), men kan bli betydligt högre om man bildar index genom att slå ihop relevanta testskalor (0.5+)
- Det är bara på det sättet man kan få en realistisk uppfattning om testens praktiska värde

# Evidensbaserade test

- Det är mycket viktigt att de test som används baseras på evidens för deras värde
- De måste med andra ord vara reliabla (fria från slumpfel i görligaste mån) och valida (faktiskt kunna användas för prognoser och förståelse)
- Testanvändare bör efterfråga evidens!
- Det är just vad STP (STP, Stiftelsen för Tillämpad psykologi) vill leverera, men ...

# Prövningar av testens värde

## - myten om "de dåliga testen"

- STP-företrädare uttalar sig mycket ofta i media och säger något i stil med "att det finns väldigt många dåliga test". Bara en handfull test (ca 5) har blivit "godkända" i deras granskningar.
- STP kräver emellertid så stora grupper vid testvalidering ( $n > 100$ ) att det blir praktiskt mycket svårt att faktiskt genomföra ett valideringsprogram. Kravet saknar rationell grund; STP kan inte motivera det.
- STP gör även en bedömning av testets värde genom att beräkna genomsnittet (medianen) av enskilda testskalors korrelationer med kriterier – detta ger en drastisk underskattning, och STP kan inte motivera sitt arbetssätt
- Som jag nyss påpekade kan relevanta analyser visa att test är mycket bättre än vad STP kommer fram till
- En utredning av testens värde genomförd av en professor i Uppsala på uppdrag av Telia för ca 15 år sedan åberopas fortfarande ofta som "bevis" för att nästan alla test måste underkännas – men utredningen är hemligstämplad och argumenten för slutsatsen kan inte bedömas.

# Problemet är inte "dåliga test", utan...

- Brist på evidens, framför allt brist på svenska studier (de flesta testerna är utländska, översatta)
- Hemlighetsfullhet och defensiv hållning inom branschen, "forskningens fiender"
- Information om test endast tillgänglig för sluten grupp, eller till stor kostnad (gäller även STP)
- Vår policy är i stället "Open Access"

Moderna personlighetstester bygger oftast på Femfaktormodellen (FFM) och "Big Five"

# Femfaktormodellen och bortom

- De fem dimensionerna är mycket allmänna:
  - extraversion
  - följsamhet
  - noggrannhet
  - öppenhet
  - emotionell stabilitet
- Aktuell forskning visar att smalare dimensioner direkt fokuserade på arbetslivet fungerar betydligt bättre
- Exempel: positiv attityd, uthållighet och kreativitet
- Index baserade på sådana dimensioner, matchade mot innehållet i det man vill predicera (kriterier), kan nå validitet omkring 0.5 – 0.6, dvs. samma nivå som begåvnings tester, givet att man korrigerar för mätfel i kriteriet och beskuren spridning i testet

# Hur långt kommer man med FFM/Big Five?

**Validitetskoefficienter (medianer) hos B5-dimensioner samt intelligens (g-faktorn) enligt meta-analys av Schmidt et al. (2008).** Schmidt, F. L., Shaffer, J. A., & Oh, I.-S. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, 61(4), 827-868.

Dimension	Arbetsprestation	Utbildningsresultat
Emotionell stabilitet	0.13	0.16
Utåtvändhet	0.09	0.17
Öppenhet	0.04	0.24
Följsamhet	0.08	0.12
Noggrannhet	0.21	0.26

# *UPP*-testet

- Nykonstruerat, svenskt test
- Grundidé: smala. Fokuserade variabler i stället för B5:s stora, abstrakta och övergripande
- Detta är ett exempel på en mycket generell psykologisk princip: "lika korrelerar med lika"
- Testet tar vid Internettetstning ca 40 minuter i fullständig version, 15 i screeningversion, och kan kombineras med begåvningsmätning som i så fall tar 16 minuter till

# *UPP* mäter förutom "Big Five": Personlighet

- Emotionell intelligens
- Social förmåga
- Samarbetsvilja
- Uthållighet
- Kreativitet
- Positiv grundhållning
- Självförtroende
- Social säkerhet
- Narcissism
- Perfektionism

# *UPP* mäter förutom "Big Five": Arbetsrelaterade attityder och motivation

- Arbetsvilja
- Arbetsintresse
- Arbetstillfredsställelse
- Balans arbete/övrigt liv
- Resultatorientering
- Förändringsvilja
- Kontrollorientering

# Normdatabaser

- Populationsdatabas (2166 testade) kan användas men har också brutits ned i delmängder:
  - Män (933)
  - Kvinnor (964)
  - 30 år eller yngre (968)
  - 31 år eller äldre (809)
  - Högskolestudier (1192)
  - Ej högskolestudier (442)
- Chefskandidater (285)
- Chefer i ledarutveckling (155)
- Norm för engelskspråkig version (120)

# Data från användning av screeningtestet (n = 6067)

- Screeningtestet har använts för 346 utannonserade tjänster
- I genomsnitt ca 30 sökande per tjänst
- De 10 vanligaste tjänstetyperna:

Kundtjänstmedarbetare
Receptionist
Säljare
Innesäljare
Ekonomiassistent
Bedrägerihandläggare
Administratör
Rekryterare
Affärsutvecklare
B2B säljare

# Hur valida är begåvnings tester?

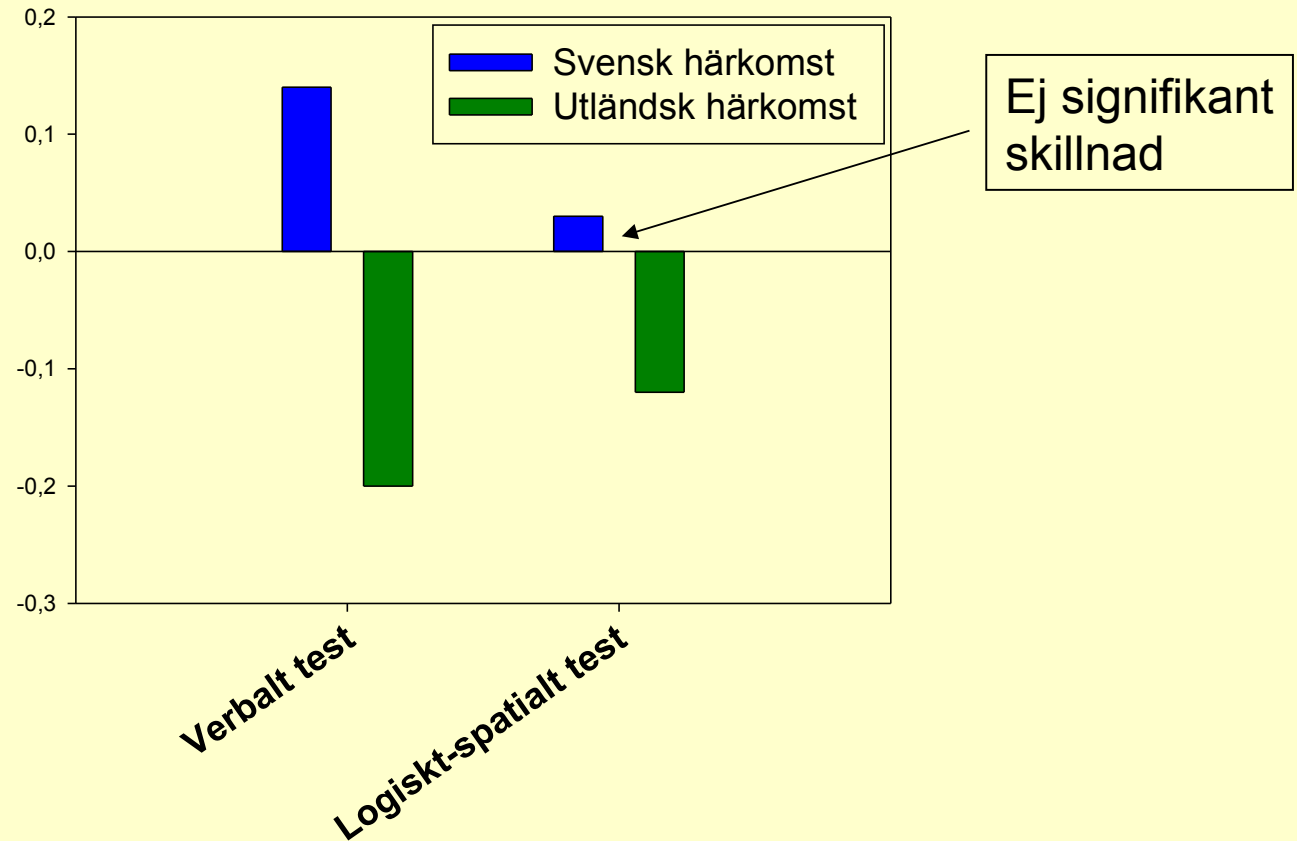
- Dessa tester är främst av betydelse för kognitiv funktion i arbetet, alltså informationsbehandling i vid mening
- Validiteten är ofta mycket hög, omkring 0.6
- Det finns många kognitiva testdimensioner, men allmänbegåvningen, *g*-faktorn, fångar oftast upp hela denna validitet
- S k praktisk intelligens har inte dokumenterats som bättre än de traditionella kognitiva testerna
- *g*-faktorn kan testas mycket snabbt (16 minuter!), men det är viktigt att skilja mellan verbala och icke-verbala (spatiala, logiska) aspekter, som kan skilja sig kraftigt för vissa grupper

# Ett snabbt test på $g$ -faktorn

- Två delprov, 8 minuter vardera:
  - Spatialt-logiskt
  - Verbalt
- Reliabilitet 0.74
- Det spatialt-logiska deltestet korrelerade 0.7 med Ravens matriser
- Det sammanslagna måttet på  $g$ -faktorn korrelerade 0.88 med Rekryteringsverkets stora I-prov efter korrektion för mätfel
- Slutsats: ett trovärdigt test av  $g$ -faktorn

# Härkomst och begåvningstestning

## Genomsnittligt testresultat (z-skala)



# Slutsats

Man kan göra en mycket snabb begåvningsstestning  
men ändå lyckas testa både  
verbala och icke-verbala aspekter separat

Detta är viktigt bl. a när det gäller mångfald  
i rekryteringen

# Hur valida är personlighetstester?

- Tidigare forskning (före 1990) tydde på att validiteten hos personlighetstester låg på maximalt 0.30 i termer av korrelation mellan test och kriterier
- Genombrottet för femfaktormodellen ("Big Five" eller B5) för personlighet omkring 1990 ledde till starkt ökad forskning och så småningom till ett antal sammanställningar av resultat (meta-analyser)
- Sambanden mellan testskalorna och arbetsresultat i skilda avseenden låg omkring 0 för 3 av dimensionerna, i intervallet 0.25-0.30 för noggrannhet/målmedvetenhet och omkring 0.20 för extraversion. Mycket omfattande data ligger bakom denna slutsats.
- Detta var egentligen ingen förbättring jämfört med tidigare resultat med andra personlighetsmodeller

Den vanliga uppfattningen att ”personlighetstester inte fungerar” är emellertid felaktig,

*men det krävs rätt testvariabler och analyser för att visa att så är fallet.*

# Exempel på valideringar mot externa kriterier

# Exempel 1: personlighet och 360-gradersbedömningar

- I aktuellt projekt i samarbete med forskare vid Karolinska Institutet har vi samlat in data på personlighet (*UPP*-testet) och 360-gradersbedömningar av chefer, kolleger och underställda i tre dimensioner: strukturering, relationsarbete och förändringsarbete
- Deltagare var chefer på mellannivå inom Stockholms Läns Landsting

**Korrelationer (validiteter) mellan 360-gradersbedömningar och matchande *UPP*-skalor efter korrektion för mätfel i kriteriet och beskuren spridning i testvariablerna ( $n = 166$ ).**

### 360-gradersbedömningar

Bedömare	Relationer	Strukturering	Förändrings- arbete
Överordnad chef	0.66	0.21	0.39
Kollegor	0.65	0.26	0.38
Medarbetare	0.66	0.40	0.51

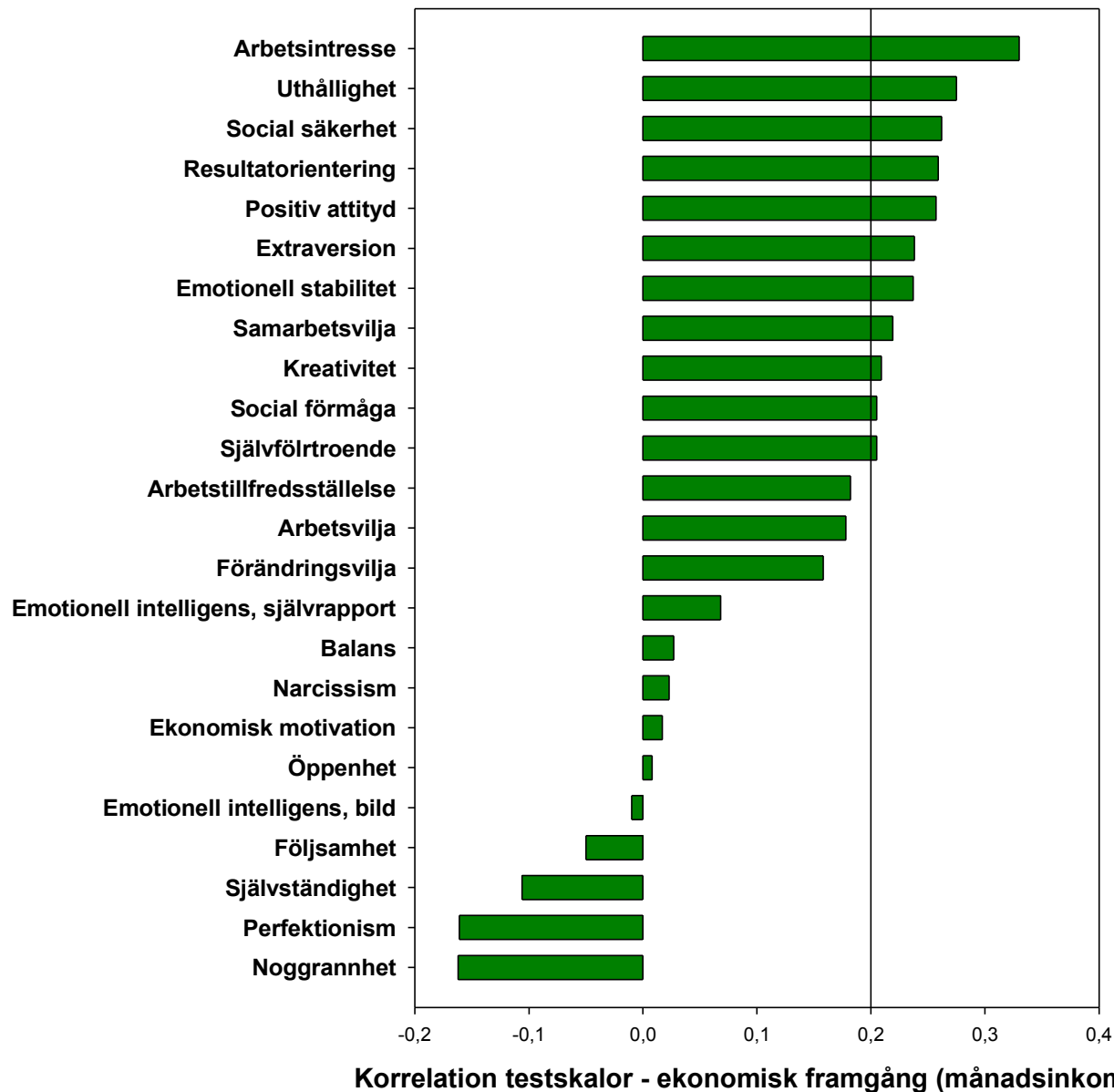
# Exempel 2: Lönekriterier vid polisen i relation till *UPP*

- Korrelation mellan *UPP* och av chefer bedömda lönekriterier ( $n = 100$ )
  - För sammanslagna lönekriterier  $r = 0.48$
  - För produktivitet  $r = 0.50$
  - För social förmåga  $r = 0.38$
  - För ett objektiva mått på arbetsresultat, antal hållna förhör under 1 år,  $r = 0.54$ .

# Exempel 3: förmansbedömningar av anställda inom kundservice i finansbolag

- Korrelationer mellan index baserade på UPP-testets skalor och kriterier:
  - Värde för företaget:  $r = 0.66$
  - Effektivitet,  $r = 0.52$
  - Social funktion,  $r = 0.40$

# Exempel 4: UPP och ekonomisk framgång



# Slutsatser: bra eller dåligt?

- Många skalor hade svaga samband med ekonomisk framgång
- Men ett index baserat på fem skalor korrelerade 0.53 med kriteriet!
- Indexet är "korsvaliderat", alltså inte ett sätt att utnyttja slump samband

Big Five-skalorna korrelerade i genomsnitt endast 0.04 med 360-gradersbedömningarna i Exempel 1, 0.02 mot polisens lönekriterier i Exempel 2

Det behövs uppenbarligen smala och fokuserade testdimensioner som matchar kriteriernas innehåll. Det är en allt vanligare uppfattning bland forskare på området.

Några av dessa kan finnas som underskalor i FFM-test, andra inte

Personlighetstest fungerar alltså,  
enligt vår och andras forskning,  
ungefär lika bra som begåvningstest.

Eftersom de har ungefär noll-korrelation  
med begåvning blir kombinationen  
mycket kraftfull.

Proxyvalidering – vad är det?

Proxy = ställföreträdare eller ersättare

# Testvalidering är en dyrbar och tidskrävande verksamhet

- Alternativ: proxyvalidering
- Mått på arbetsprestation är mycket svåra att samla in, dels i princip dels för att företagen oftast är ovilliga att medverka till det
- Proxy-kriterier samlas enkelt in i samband med testningen
- Forskning finns som stödjer antagandet att de fungerar som indirekta mått på arbetsprestation
- De har även ett värde i sig och underlättar starkt allt valideringsarbete

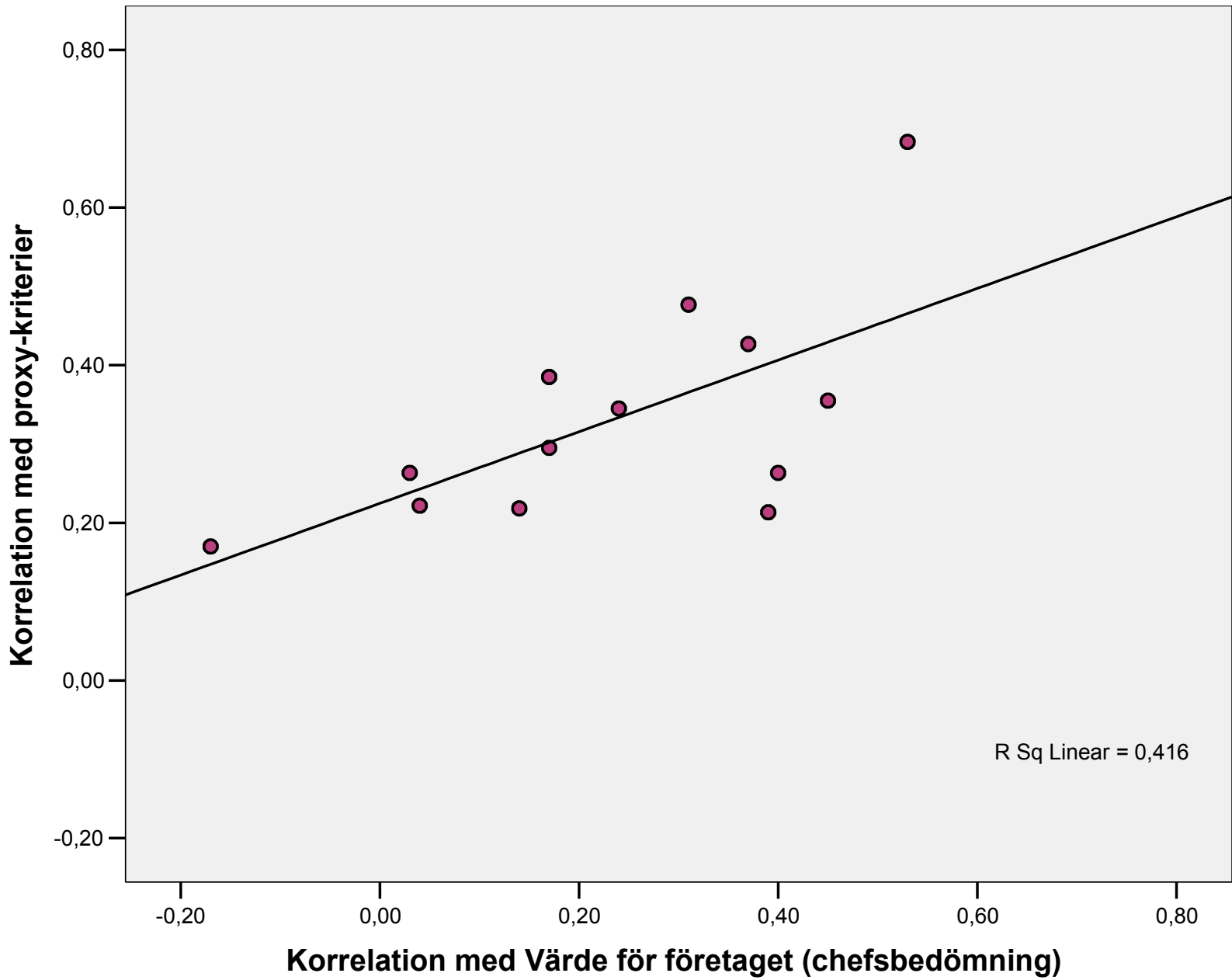
# Arbetsrelaterade dimensioner ”Proxy-kriterier”

- Arbetsvilja
- Arbetstillfredsställelse
- Arbetsintresse
- Förändringsvilja
- Resultatorientering
- Kontrollorientering
- Balans arbete – övrigt liv

# Korrelationer mellan proxykriterier och ett objektiva, kvantitativt mått på arbetsresultat (antal genomförda förhör under ett år, polisstudie)

Arbetsstillfredsställelse	0.37
Arbetsvilja	0.44
Arbetsintresse	0.37

Kan vi verkligen tro att proxykriterier ger information som är jämförbar med externa kriterier på arbetsresultat?



## Validering mot 6 proxykriterier, multipla korrelationer (kvadrerade)

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

# Slutsats

- I flera studier har proxyvalidering visat sig ge resultat som är approximativt likvärdiga med extern validering, möjligen något lägre
- Vi rekommenderar proxyvalidering som en enkel och kostnadseffektiv metod att validera test och andra typer av information
- Hela *UPP* behöver i så fall inte användas, det räcker med den relevanta modulen

# Så långt är allt gott och väl, men ...

- Personlighetstester är oftast av självrapporttyp
- Risken är stor i sådana tester att många testade personer beskriver sig själva överdrivet positivt, att de skönmålar
- och därigenom vinner en orättvis fördel
- Vad göra? Ipsativt svarsformat?

## Ipsativa test (mycket vanliga, tex OPQ, MPA och "Thomassystemet")

- Avser att korrigeras för skönmålning men lyckas inte – effekten finns kvar
- Jättestora problem vid fastställande av psykometriska egenskaper
- Tar längre tid och är inte omtyckta av de testade som anser att de instrueras att "jämföra äpplen och päron"
- Ger inte data som är jämförbara *mellan* personer – katastrofalt vid urval!
- Ger inte heller rättvisande data *inom* personer – "gummisnoddseffekt"

# Gumminsuddseffekten vid ipsativ testning

- 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 - fullständigt vilseledande.
- Det här är inte en teoretisk spekulation utan något jag sett i praktiken.
- Man kan bara spekulera i hur många felaktiga beslut som har fattats under årens gång på grundval av dessa ipsativa test.
- Test som tillmäts mycket stor betydelse i många sammanhang!

# Fallstudie

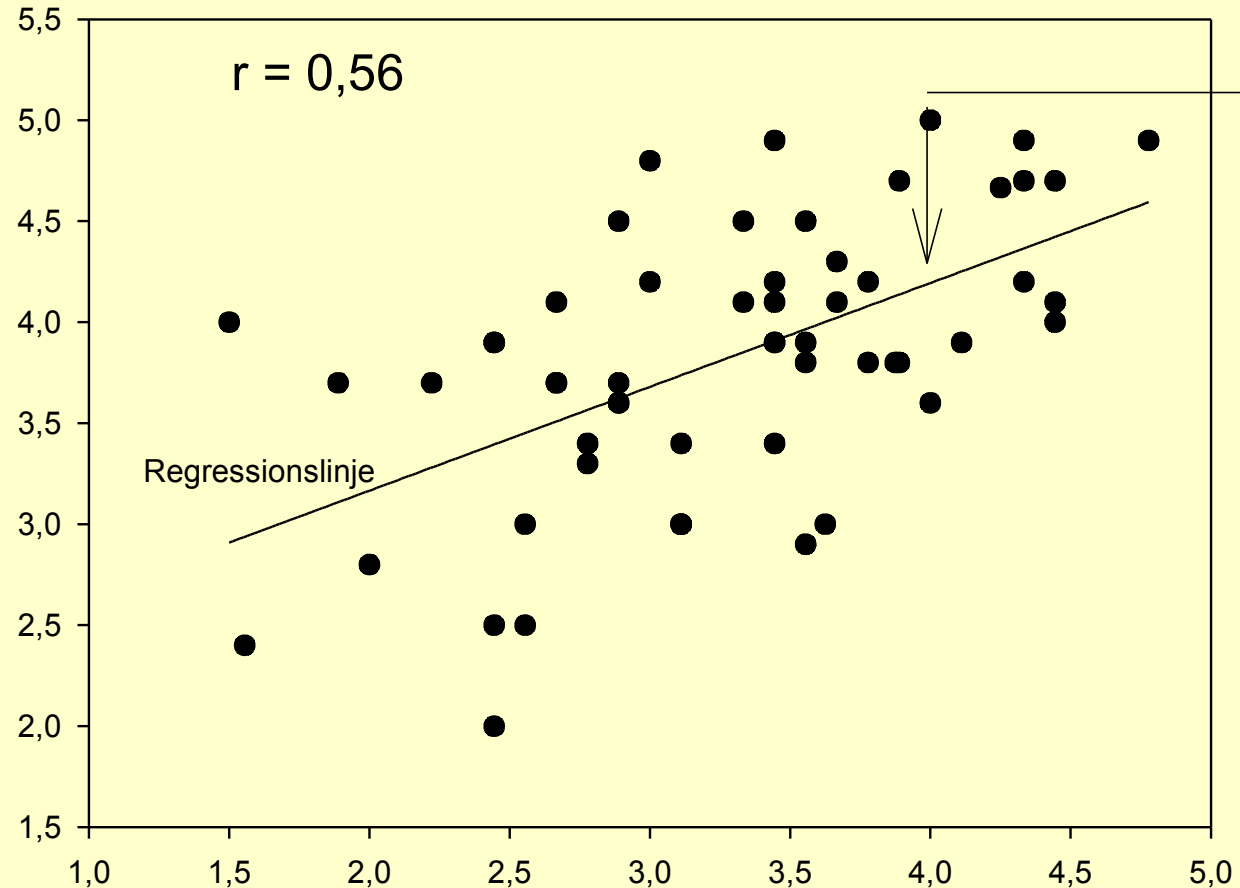
- X var kandidat till ett viktigt chefsjobb vid en stor bank
- Han hade goda resultat vad gällde referenser, i ett flertal intervjuer och även en medarbetarundersökning vid sitt tidigare chefsjobb
- Han OPQ-testades och fick lågt värde i extraversion – HR-avdelningen avrådde starkt från anställning enbart på den grunden
- Search-konsulten, som vi samarbetar med, bad mig ordna en testning med *UPP* och intervju
- X visade sig ligga högt i många av *UPP*-skalorna, men relativt sett lägre i extraversion (6 på skalan 1-9)
- OPQ:s gummisnoddmätning hade gett honom ett lågt värde helt enkelt därför att han var bättre i andra avseenden
- Det slutade med att han trots allt blev anställd

# Korrektion med hjälp av särskilda skalor som mäter tendensen att svara socialt önskvärt

- I testet inkluderas en eller flera skalor som mäter den testade personens tendens att svara vad han eller hon tror är socialt önskvärt
- Sådana skalor har visat sig vara starkt korrelerade med vissa (men inte alla) personlighetsskalor
- Effekten av skönmålning kan elimineras med statistisk metodik

# Korrektion för skönmålning

Samarbetsvilja (omvändningen av passiv aggression)



Korrigerade värden  
mäts som residualer

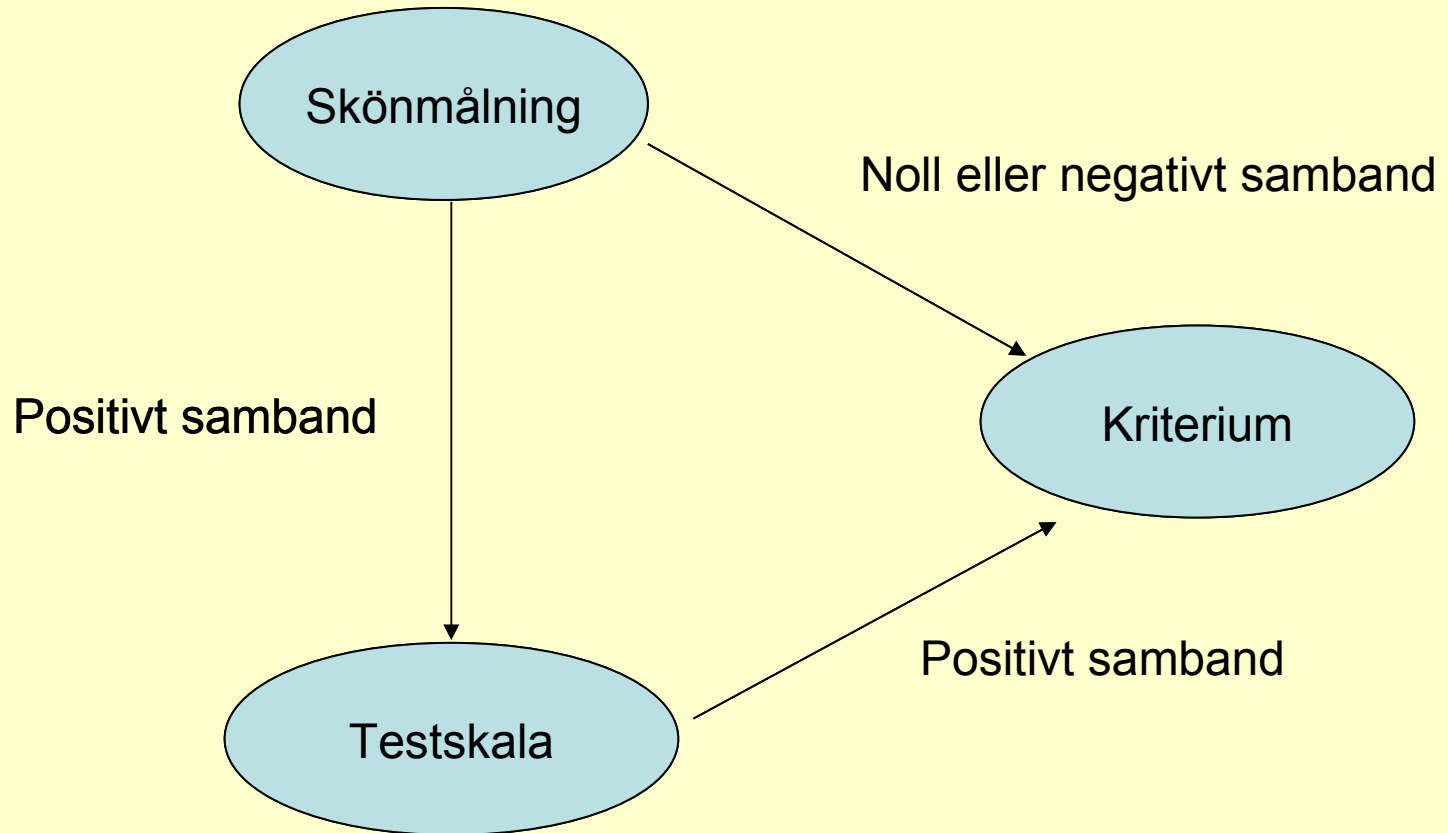
## Viktigt: SPECIFIK korrektion

Testvariabel	Andel av variansen som förklaras av skönmålning (två skalor)
Extraversion	0,139
Vänlighet	0,258
Emotionell stabilitet	0,276
Öppenhet	0,104
Noggrannhet	0,098
Uthållighet	0,112
Samarbetsvilja	0,243
Positiv grundattityd	0,338
Självförtroende	0,273
Social förmåga	0,153
Emotionell intelligens	0,207
Kreativitet	0,023
Perfektionism	0,042

# Kritik och svar

- IM (impression management) är egentligen en personlighetskala
- Svar: sant att den har en viss generalitet, men det betyder inte att IM inte mäter svars-bias i en enskild testning
- IM är positivt korrelerad med social anpassning och liknande dimensioner, som ofta är kriterier i testvalidering
- Svar: detta är sant ibland, men korrelationen kan vara svag eller negativ – i så fall fungerar IM som vi hoppats som en ”suppressor”, och validiteten ökar genom korrektion för IM
- Dessutom elimineras effekter av ”faking” både i experimentella studier och fältstudier
- Om man INTE korrigerar kvarstår enorma effekter av ”faking” och de som bluffar gynnas storligen

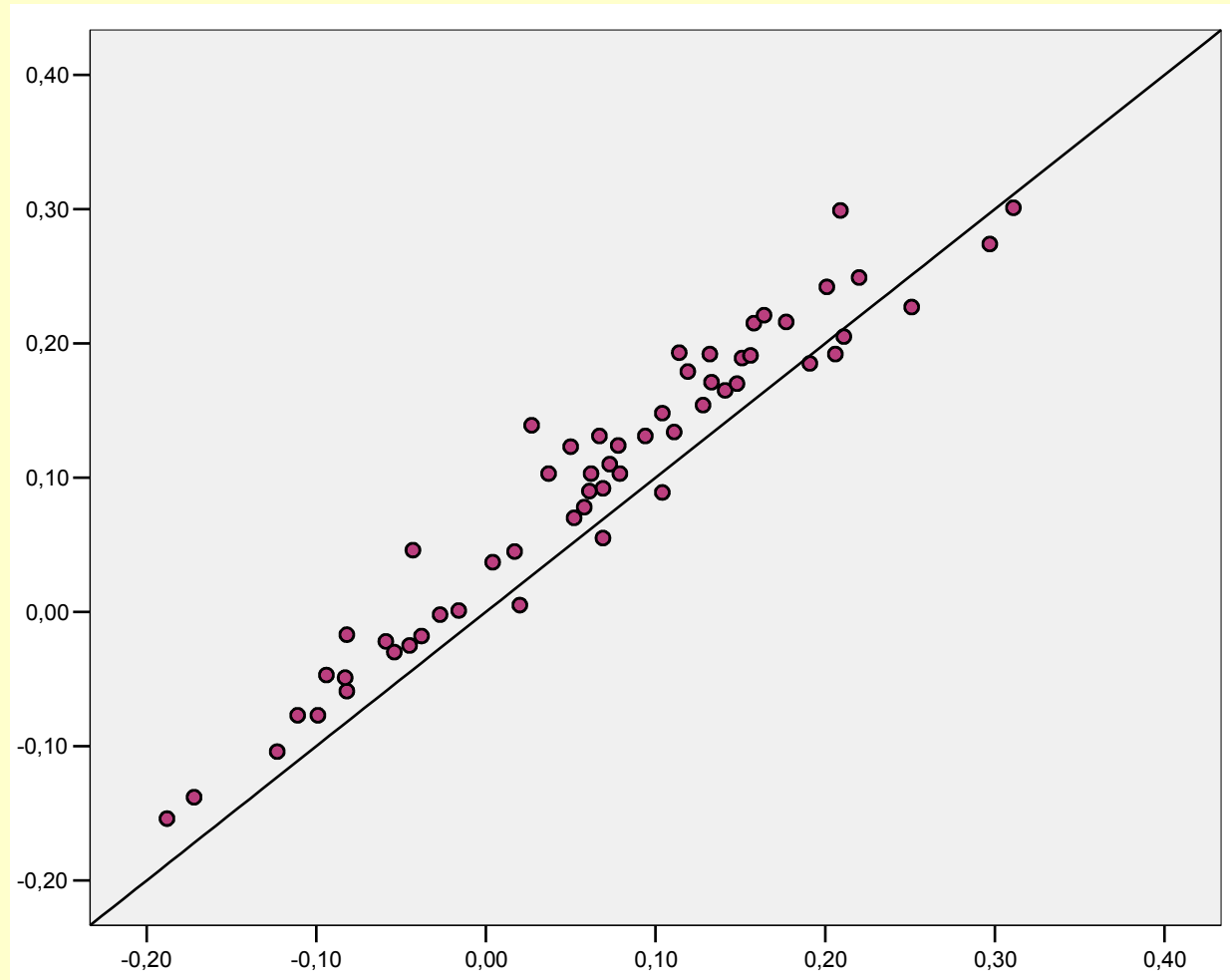
# Skönmålning som suppressor



## Korrelationer mellan 360-gradersbedömningar och skönmålning.

	Overt skönmålning	Kovert skönmålning
Struktur, bedömd av chef	-0.04	0.16
Relationer, bedömd av chef	-0.02	0.22
Förändring, bedömd av chef	-0.08	0.20
Struktur, bedömd av	-0.16	-0.17
Relationer, bedömd av kollegor	-0.17	-0.08
Förändring, bedömd av kollegor	-0.22	-0.02
Struktur, bedömd av underställda	-0.11	-0.03
Relationer, bedömd av underställda	-0.09	-0.00
Förändring, bedömd av underställda	-0.17	0.01

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



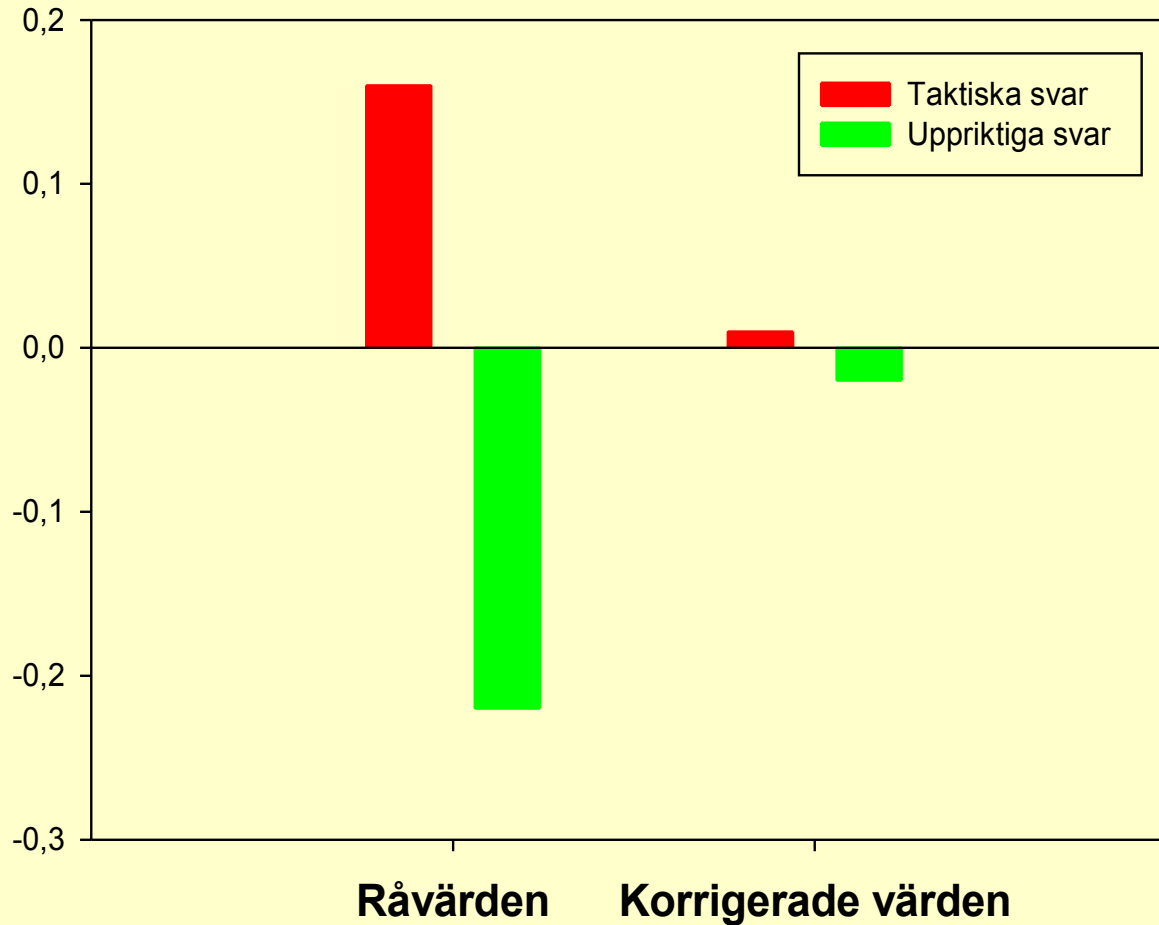
# Med andra ord...

- Korrektion för skönmålning ökar testskalornas validitet
- En suppressoreffekt!

# Experimentell studie av "faking"

## Personlighetsskalor

Standardiserade testpoäng



Ca 90 % av effekten av skönmålning eliminerades

# Skarpt läge vs normdata, screeningtestet *UPP/Screen* Extraversion

Standardiserad testskala



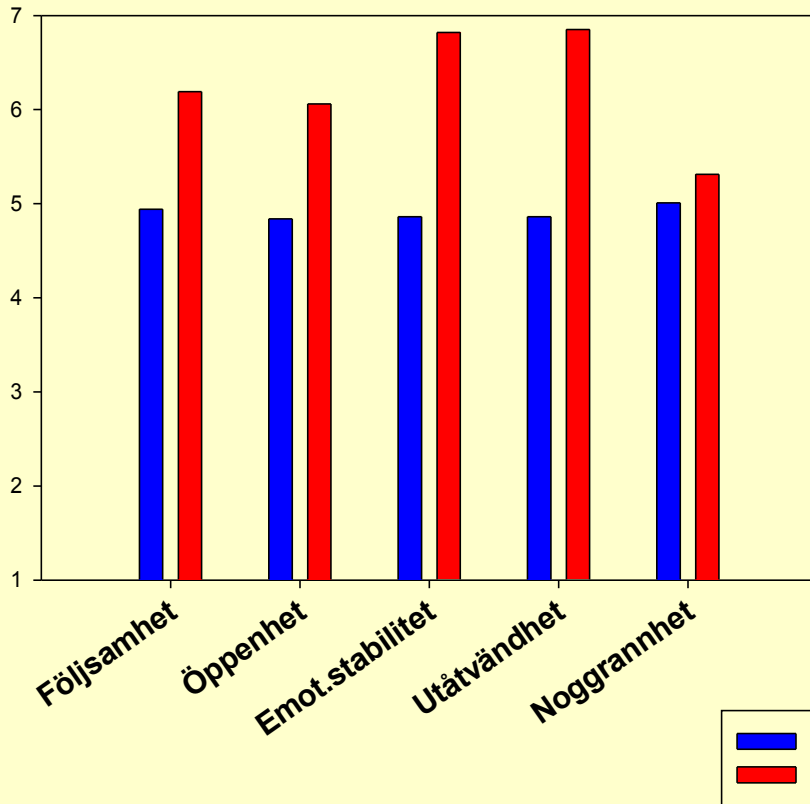
**Före korrektion skillnaden  $p < 0,00005$ ,  
efter korrektion ej signifikant**

# Personlighetstestning vid chefsrekrytering

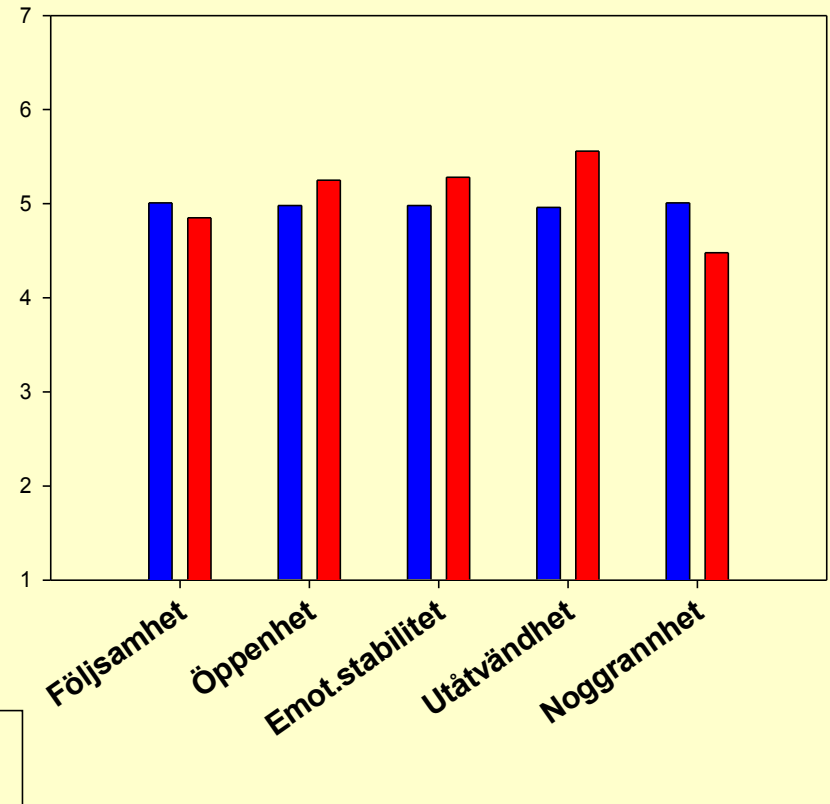
# Chefskandidater skönmålar!

## Medelvärden i testvariabler, stanine

Före korrektion

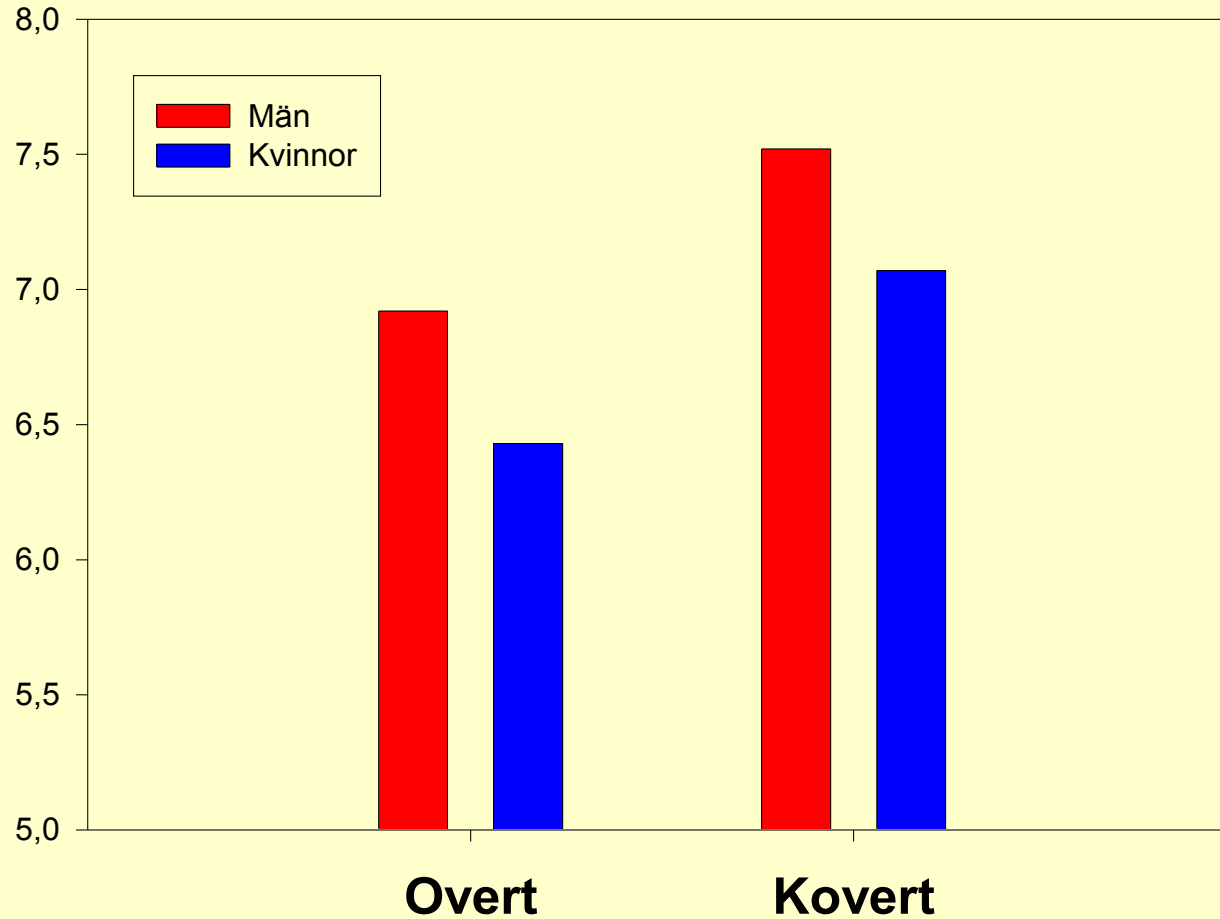


Efter korrektion



# Skönmålning hos chefskandidater

## Medelvärde



Hög skönmålning hos alla, men högst hos män

Overt=direkt mått, typ social önskvärdhet

Kovert=indirekt mått

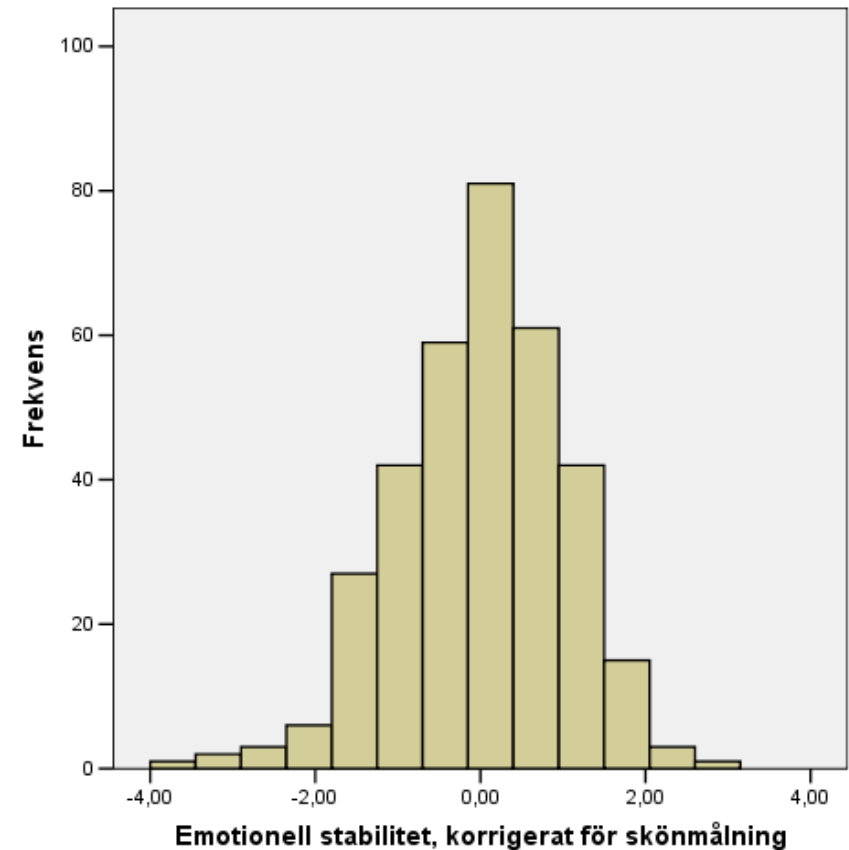
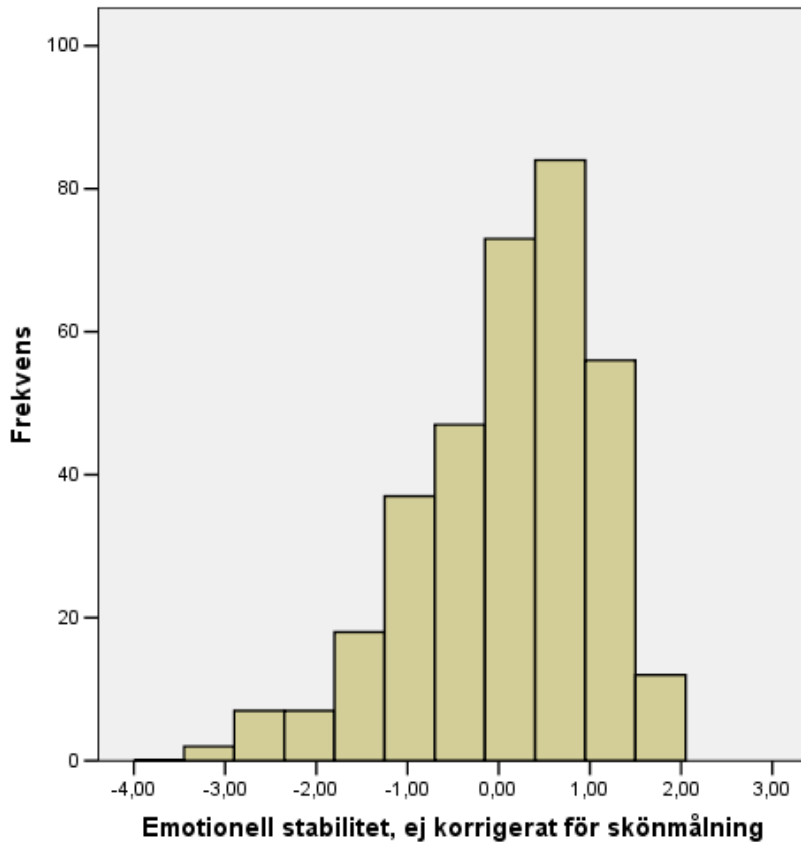
Skala: M=5, SD=2 i normgruppen

Lennart Sjöberg

58

AOnet 11-09-27

# ”Takeffekt” i okorrigerad version av Emotionell stabilitet



# Konsekvenser

- Om man inte korrigerar för skönmålning missgynnas kvinnliga sökande till chefstjänster
- Om man korrigerar försvinner effekten av skönmålning – de som ”bluffar” gynnas inte längre
- Om man inte korrigerar tenderar testresultaten vid chefsrekrytering att ”slå i taket” och vara ganska oanvändbara

# Jämförelser av testresultat för jobbsökande med svensk eller utländsk härkomst

## Använt test:

Screeningversionen av *UPP*-testet, som mäter

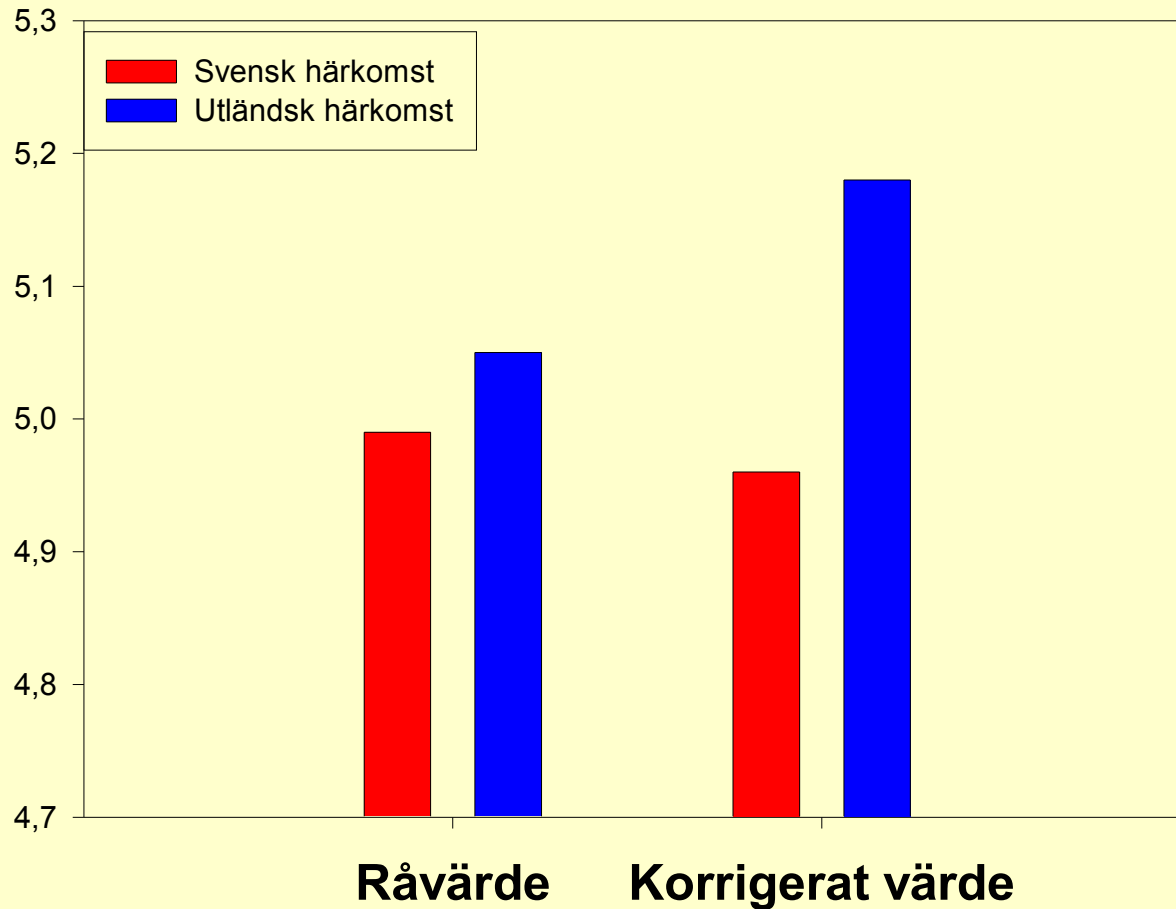
- Utåtvändhet
- Samarbetsvilja
- Kreativitet
- Uthållighet
- Positiv hållning

samt

- Tendensen att skönmåla

# Positiv hållning

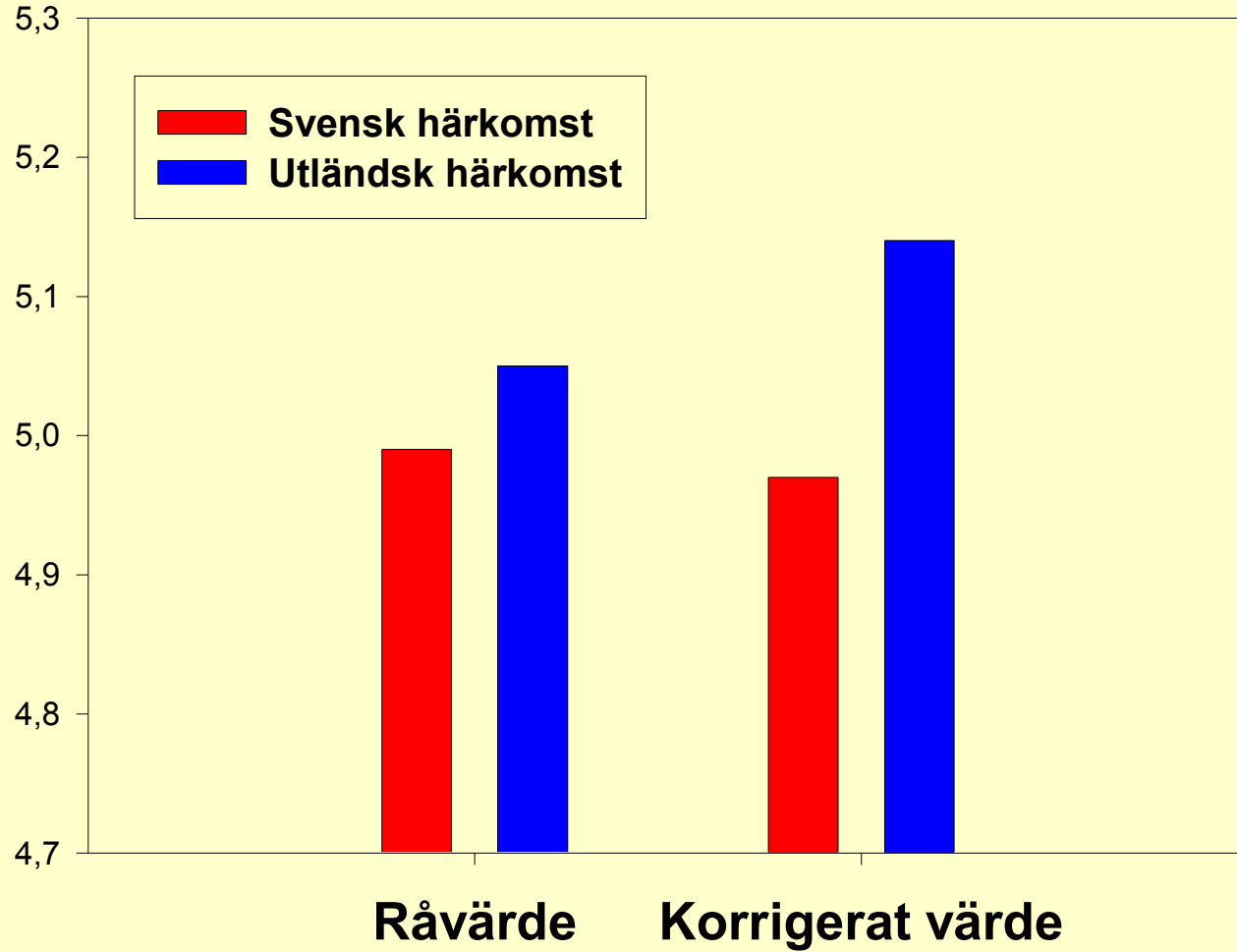
## Medelvärde



Skala: M=5, SD=2

# Utåtvändhet

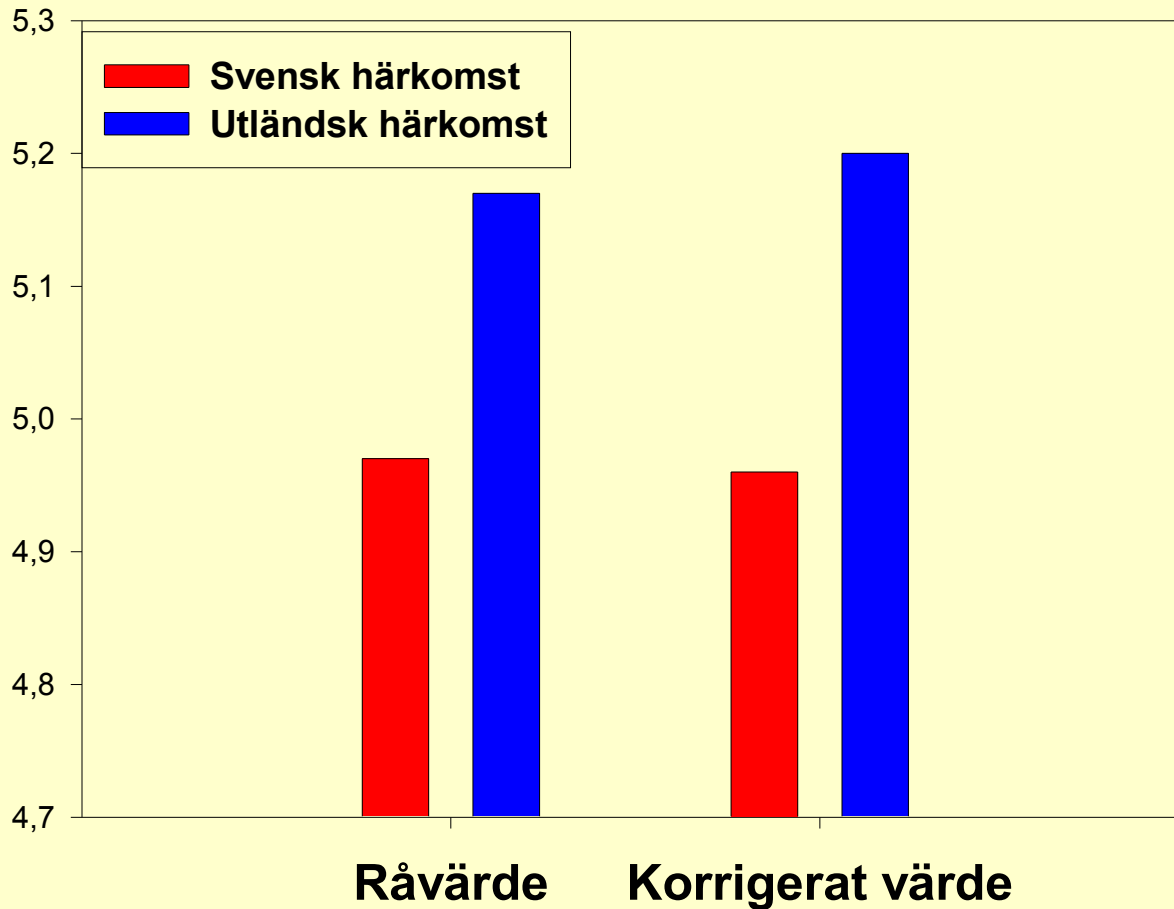
## Medelvärde



Skala: M=5, SD=2

# Kreativitet

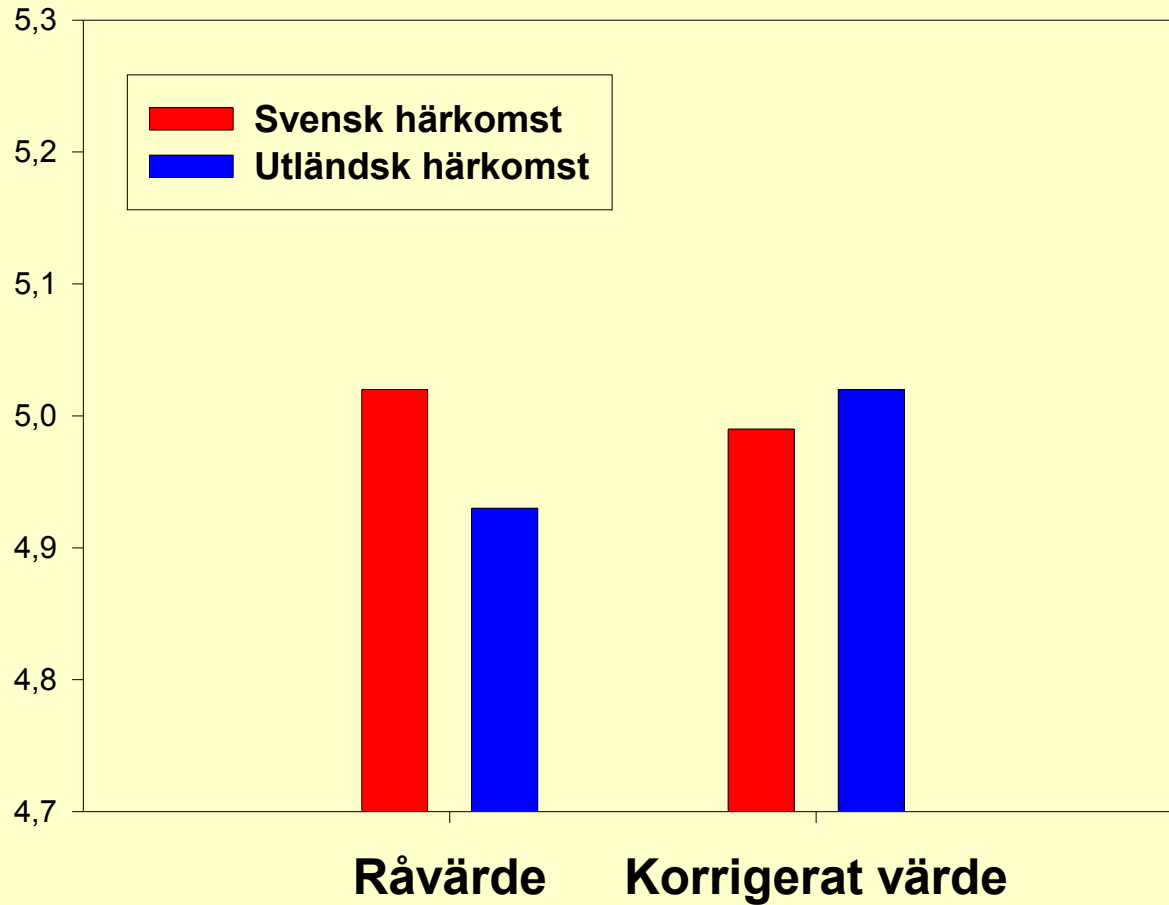
## Medelvärde



Skala: M=5, SD=2

# Uthållighet

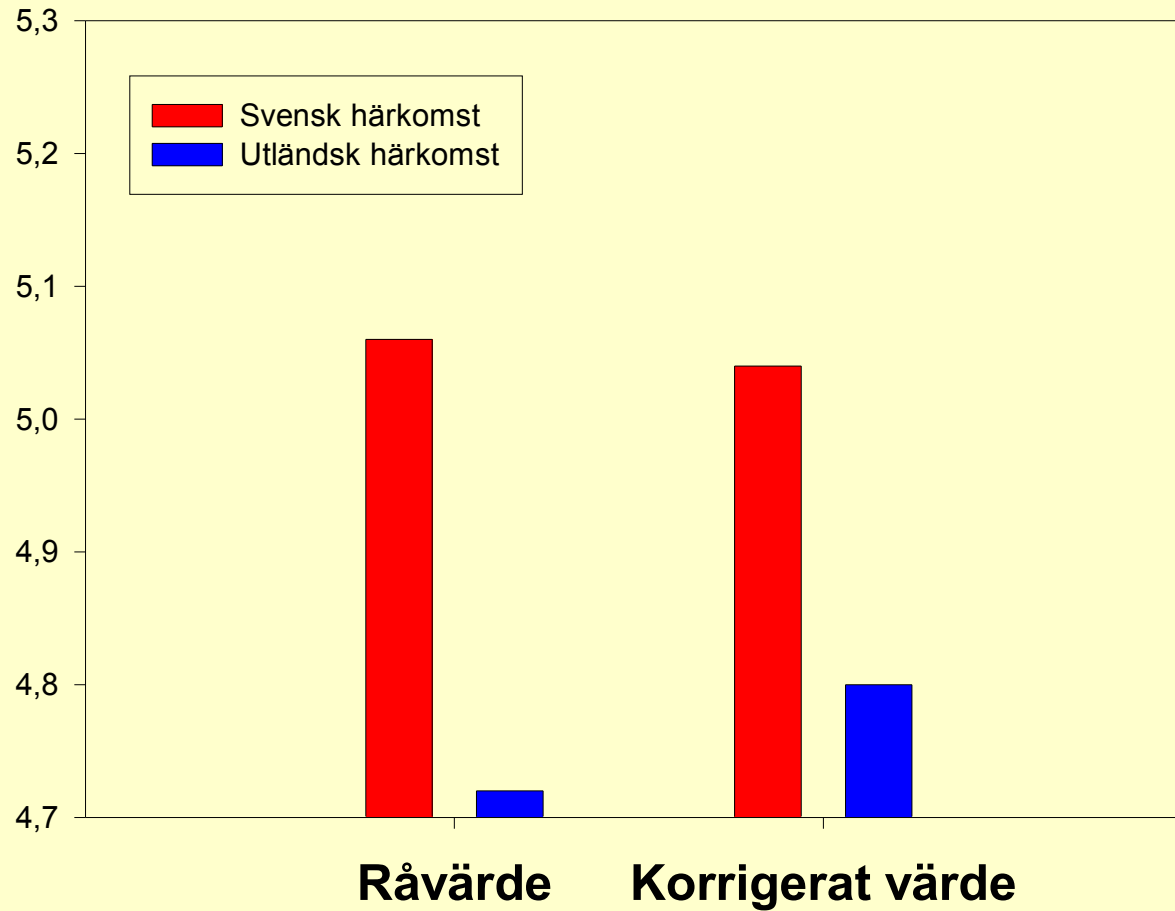
## Medelvärde



Skala: M=5, SD=2

# Samarbetsvilja

## Medelvärde



Skala: M=5, SD=2

# Vad betyder dessa skillnader i konkurrens om jobben, om man tar hänsyn till testresultaten?

- I genomsnitt var det 30 sökande/jobb, alltså var det bara 3.3 % som kunde anställas
- Om de fem testskalorna sammanvägdes med lika vikt till ett "lämplighetsindex" var det 2.7 % av dem med utländsk härkomst som hamnade gruppen av de 3 % som hade högsta värdet
- Efter korrektion för skönmålning steg den siffran till 4.1 %
- Chansen att hamna på "valbar plats" ökade alltså med **152 %** för dem med utländsk härkomst, om korrektion gjordes för skönmålning

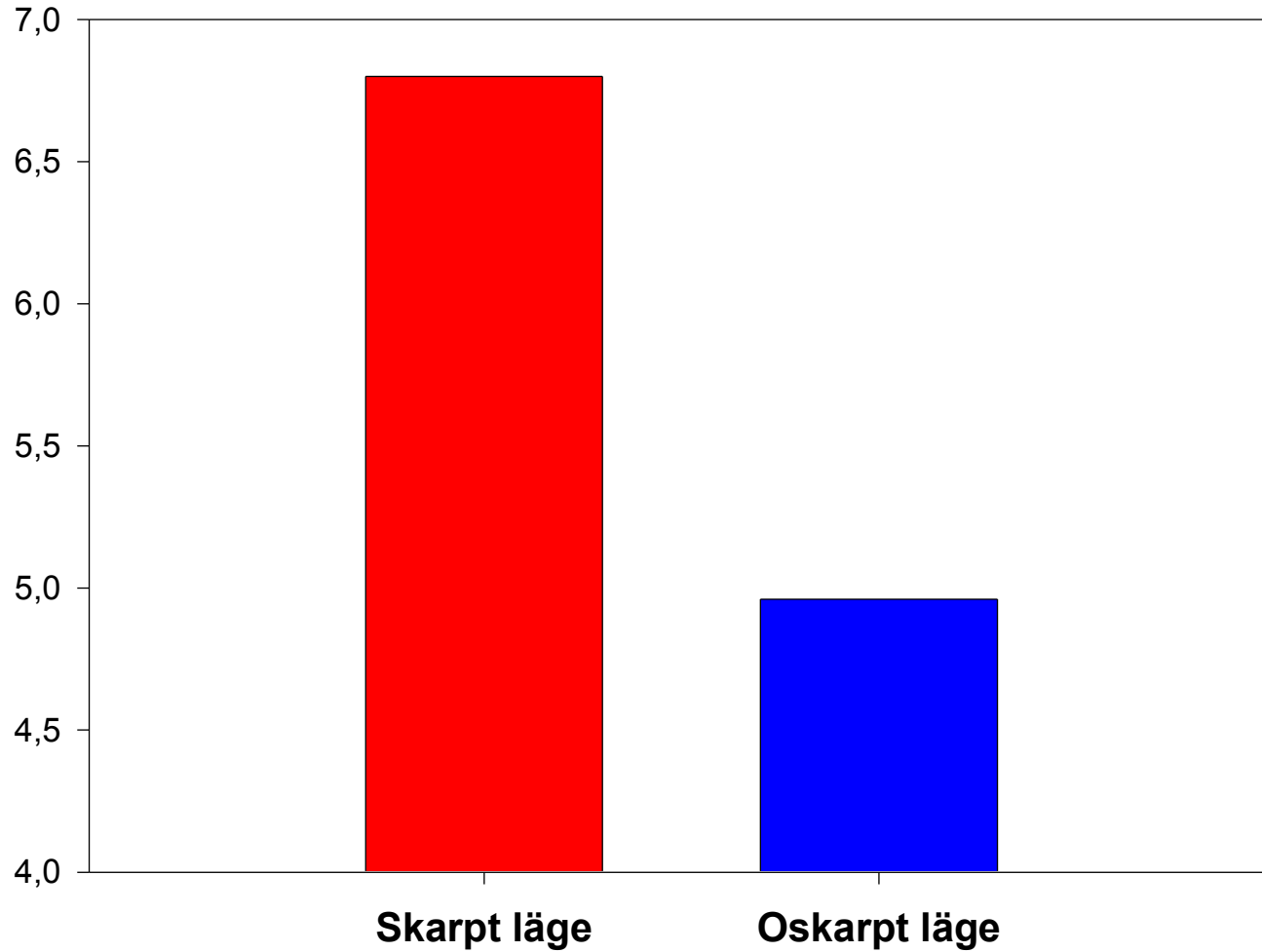
# Slutsatser om härkomst

- Jobsökande med utländsk härkomst hade högst medelvärde i 4 av 5 dimensioner
- Skillnaden utländsk – svensk härkomst ökade efter korrektion för skön målning; de med utländsk härkomst skönmålade mindre
- Chansen att hamna i tätgruppen ökade med 152 % för dem med utländsk härkomst, efter korrektion – gynnar mångfald!

# Datakvalitet

- Alla testade är inte högt motiverade för uppgiften att besvara testuppgifterna
- Låg motivation ger låg datakvalitet och eventuellt missvisande resultat
- Mätning av datakvalitet:
  - Intra-individuell spridning av svaren
  - Ja-sägartendens
  - Strukturell likhet mellan en persons svar och normdata
  - Sammanvägs till ett mått på datakvalitet

# Datakvalitet



Skala: M=5, SD=2

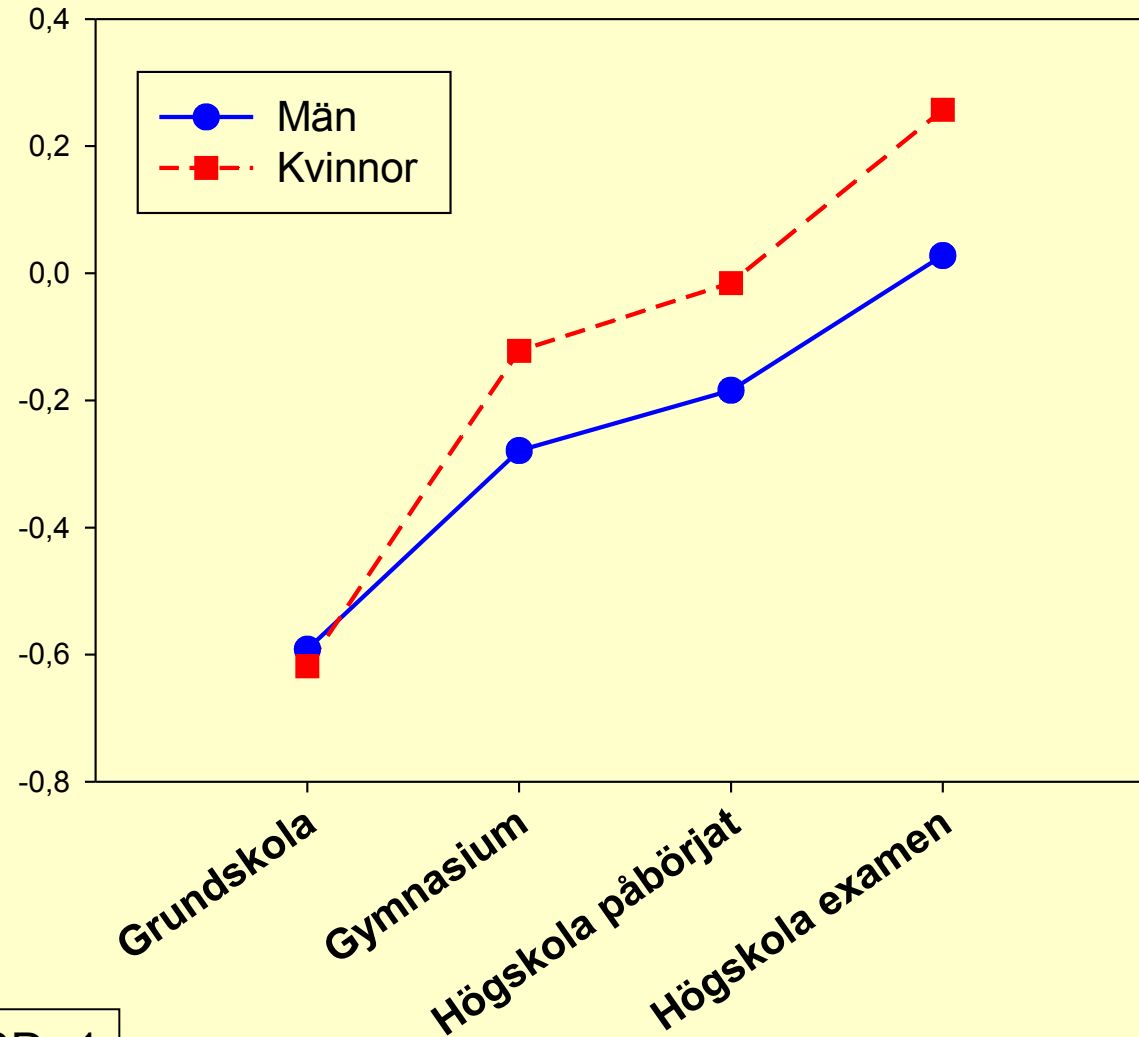
Lennart Sjöberg

70

AOnet 11-09-27

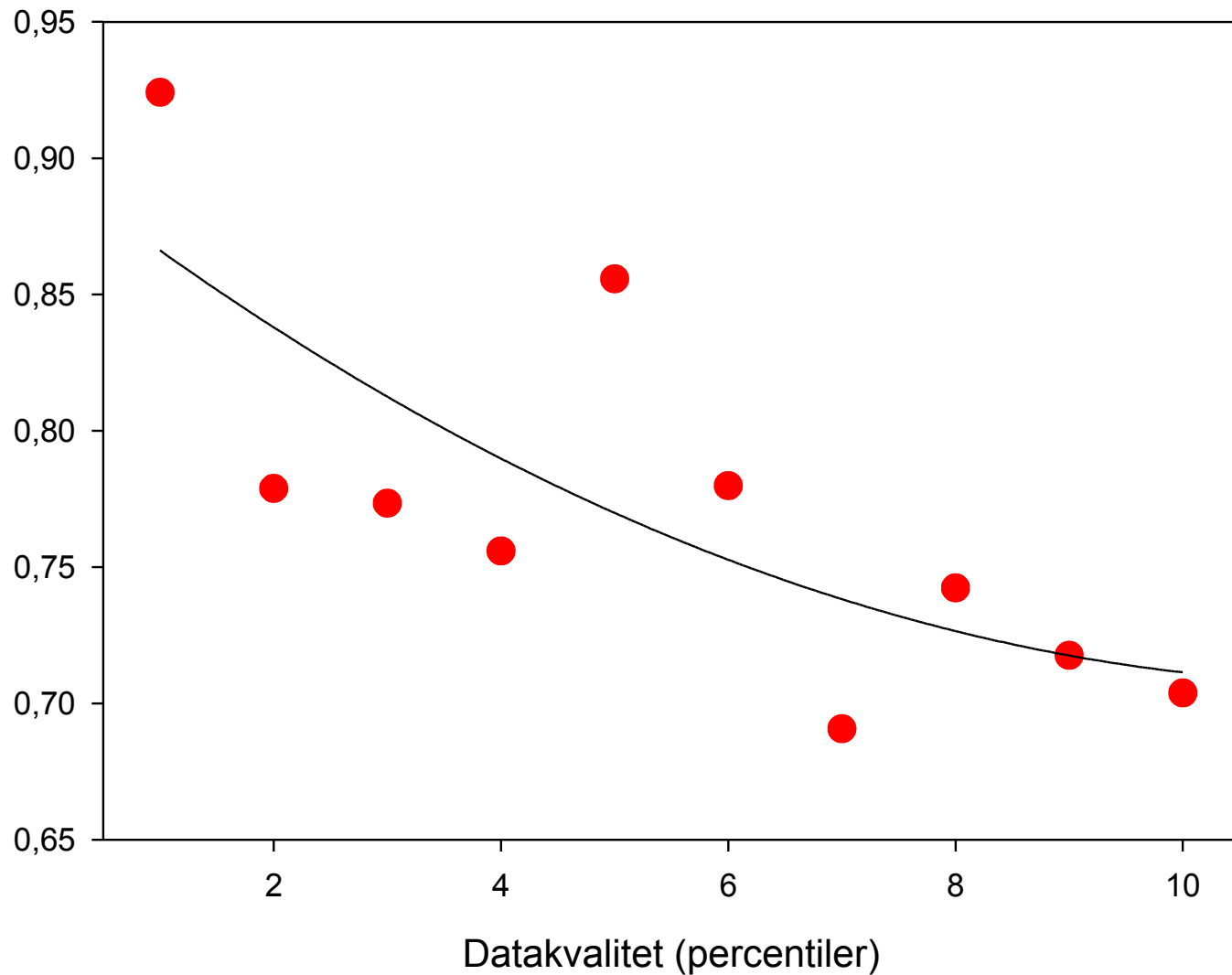
# Datakvalitet, kön och utbildning

## Medelvärde i datakvalitet



Skala: M=0, SD=1

# Genomsnittligt prognosfel



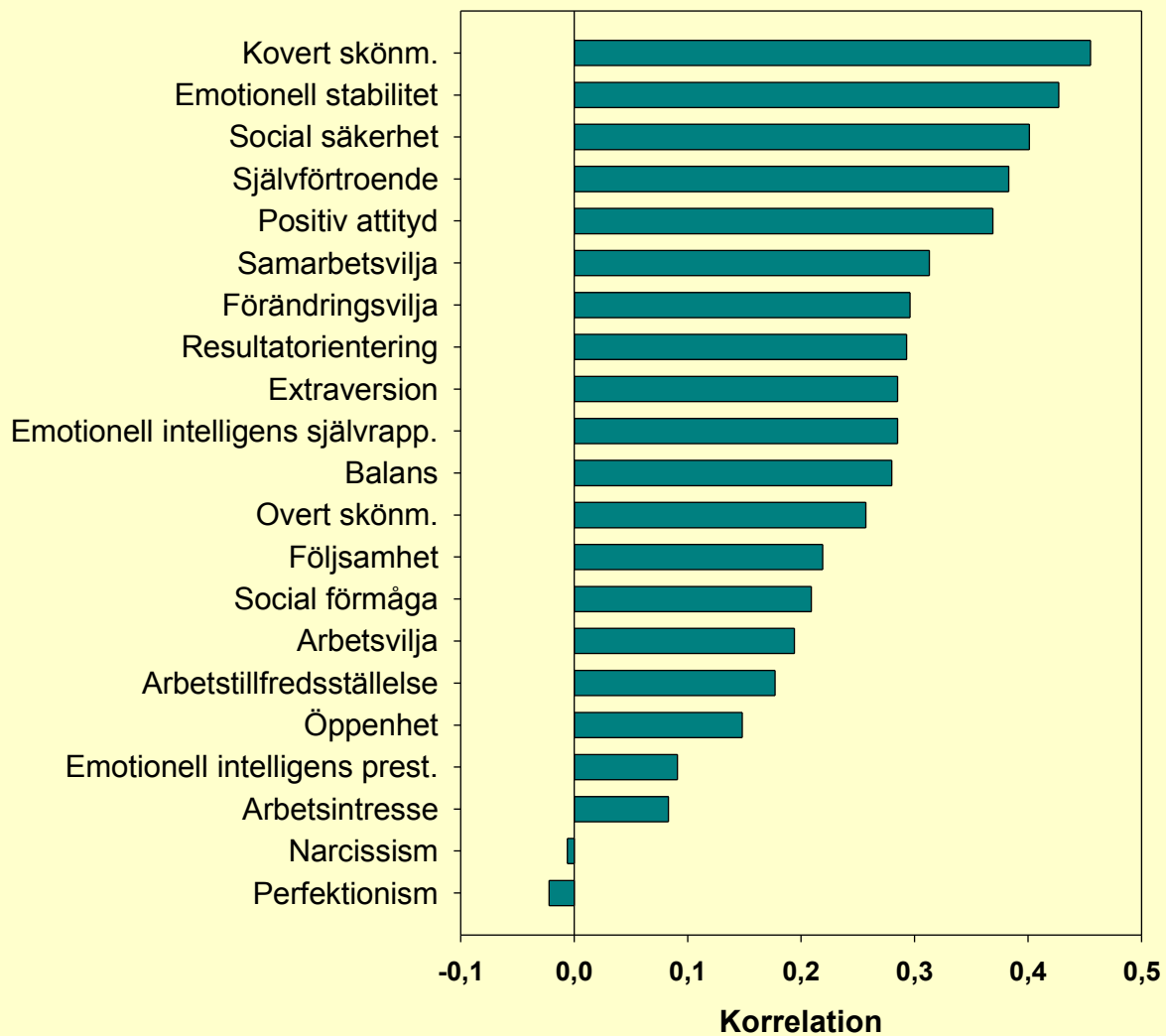
# Emotionella och värderande reaktioner på testning

- Sällan eller aldrig beaktade vid testningar
- Kan ändå vara av stor betydelse för tolkning av resultaten
- Vi tar därför alltid in data på stämningsläge och attityd till testet
- Vad säger sådana data?

# Mätning av stämningssläge

- Stämningsslägets tre dimensioner:
  - Spänd/lugn
  - Trött/pigg
  - Ledsen/glad
- Dessa tre dimensioner mäts reliabelt med sammanlagt 12 frågor i början av testningen
- De slogs ihop till ett index: *Humör*

# Korrelation mellan humör och personlighetsskalor



# Kommentarer

- Ganska starka samband mellan personlighetsskalorna och stämningläget vid testningens början
- Det betyder att den som var "ur slag" vid testningen kan ha fått missvisande resultat
- Viktigt att veta vid tolkningen och annan användning av testresultatet om stämningläget var onormalt, sedan eventuellt omtesta
- Kvinnorna hade lägre stämningläge (mera trötta, spända och nedstämda) än männen, signifikant skillnad ( $p = 0.001$ ). Storleken på effekten var ca  $\frac{1}{4}$  standardavvikelse.
- Kvinnor kan missgynnas därför om man saknar information om denna faktor

# Attityd till testet och testningen

- Mäts med 7 frågor i slutet av testsessionen
- Dessa kombineras till ett index som varierar från 1 (lågt betyg för testet) till 5 (högt betyg).
- Viktigt för trovärdigheten hos testanvändningen!
- Okänt för de flesta tester

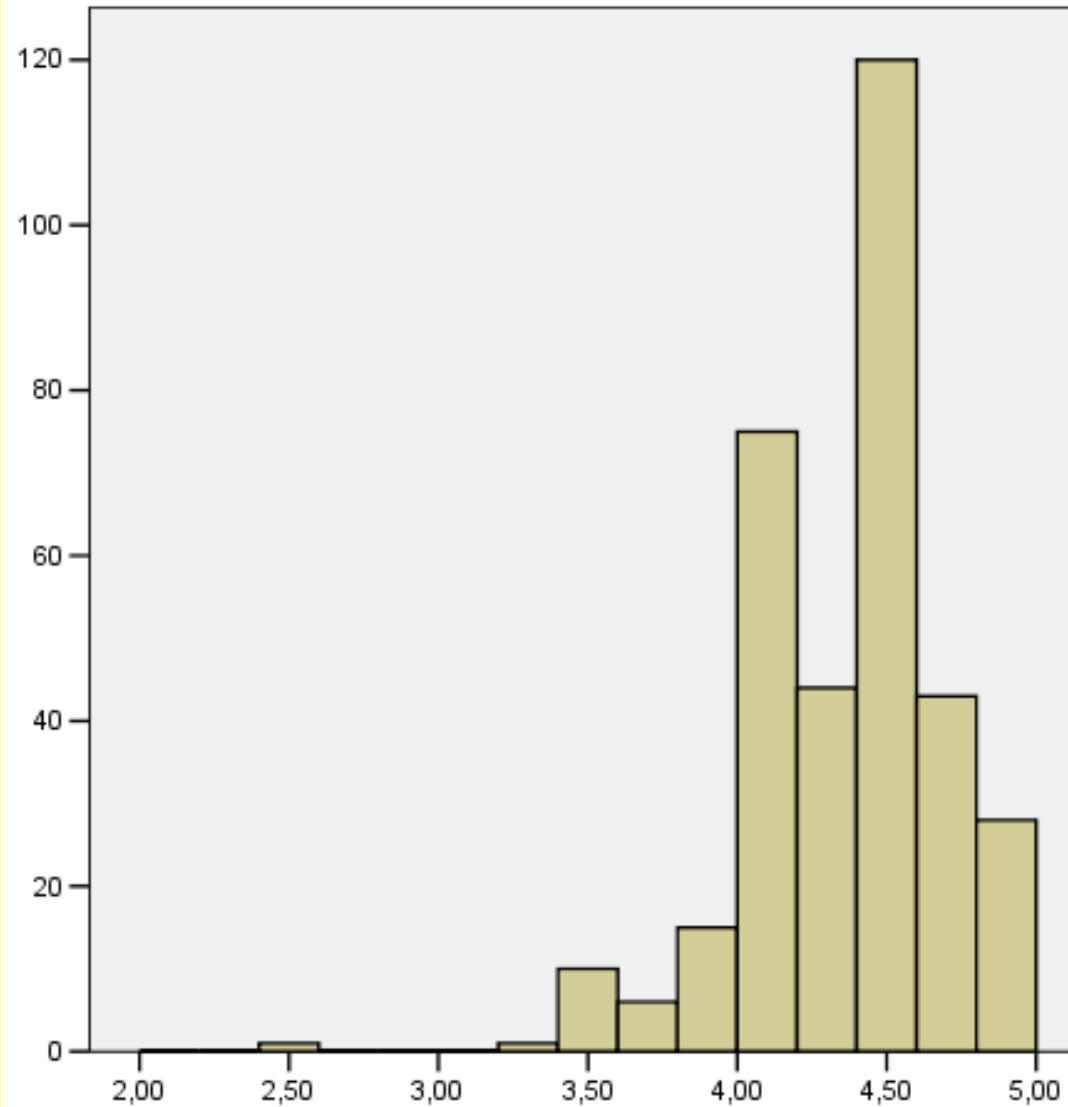
# Syfte med attitydmätningen

- Kvalitetskontroll av hur de testade upplever testet
- Viktigt för trovärdigheten inte bara av själva testet utan av hela processen
- Förekommer bara i *UPP*-testet, såvitt bekant

# Attitydskalan

- Anser du att testningen var meningsfull?
- Tog testet upp viktiga frågor?
- Var instruktionerna klart formulerade?
- Var testuppgifterna klart formulerade?
- Var svarsalternativen klart formulerade?
- Var det en intressant uppgift att besvara testuppgifterna?
- Var det något som saknades i testet?
- Var testet bättre eller sämre än andra personlighetstest som du har tagit?

# Frekvens



Genomgående  
positiva  
bedömningar  
av testet

Negativ

Attityd till testet

Positiv

Lennart Sjöberg

80

AOnet 11-09-27

# Ny version: *UPP 2.0*

- Flera skalor, med möjlighet att bara använda de skalor man har intresse av i varje särskilt fall
- Nya rapportfunktioner
- Helt självadminstrerande plattform

# Utveckling av IT-metodiken

- Testning på Internet ekonomiskt lönsam och ger jämförbara resultat med traditionell testning
- Undantag kan vara begåvningstest som kan kräva möjlighet till övervakad verifieringstestning
- Automatisk rapportgenerering önskvärd (tidsfaktorn!), olika rapportformat ska finnas och vara valbara
- Olika normdatabaser skapas när tillräckliga underlag finns och ska sedan vara valbara. Ny forskning visar att drygt 100 är tillräckligt för att ge en stabil norm
- Krypterad dataöverföring, helt inom landet. Observera integritetsproblem vid dataöverföring till utlandet!

# Slutsatser

- Tester kan fungera bra för både begåvning och personlighet (undantag finns)
- Validitetssiffrorna ligger omkring 0.5 – 0.6 *för smala, fokuserade personlighetsskalor*, inte för "Big Five"
- Vid användning av självrapporttester är det nödvändigt – och möjligt – att korrigera för skönmålning
- Stämningläge, datakvalitet och attityd bör också mätas för att kvalitetssäkra testresultaten
- IT-metodiken möjliggör numera snabba och säkra testningar (så länge data stannar inom landet)
- Kvalitetssäkrade tester gynnar mångfald och jämställdhet

# Mera information på

- [www.psykologisk-metod.se](http://www.psykologisk-metod.se) bl a manualer och forskningsrapporter för gratis nerladdning, tillgängliga för alla
- eller skriv till [lennartsjoberg@gmail.com](mailto:lennartsjoberg@gmail.com)