

# Ipsativa och normativa svarsformat i personlighetstest

**Lennart Sjöberg**

**Handelshögskolan i Stockholm**

## Sammanfattning

Skönmålning är ett stort problem vid användning av personlighetstest i självrapportformat. Ett vanligt sätt att försöka lösa detta problem är att använda jämförande, eller ipsativa, svarsformat. Den ansatsen har många problem. Dels innebär den besvärliga psykometriska problem, dels upplevs den som negativ av de testade. Ipsativa format bygger på jämförelser *inom* individer och kan logiskt sett inte användas för jämförelser *mellan* individer, något som gör dem i princip oanvändbara vid urval. Forskning visar dessutom att ipsativa format inte lyckas eliminera effekterna av skönmålning. Det kan man emellertid göra till åtminstone 90 % med ett normativt svarsformat och användning av särskilda skalor för att mäta social önskvärdhet. Dessa skalor används sedan i statistiska modeller för korrektion. Eftersom kvinnor och invandrare skönmålar mindre än andra grupper missgynnas de om man inte korregerar.

Många test som används i arbetslivet syftar till att mäta personligheten med uppgifter av självrapporttyp. Så här kan en typisk uppgift se ut:

*Jag går gärna på fester*

Håller absolut med   Håller med   Tveksam   Håller inte med   Håller absolut inte med

Detta är ett normativt format (Likertformat). Det är lätt att handskas med i praktiskt arbete med att konstruera och använda test. Den statistiska analysen är rättfram. Vi sätter helt enkelt siffrorna 1-5 på svarsalternativen och behandlar dem som om de vore mått på en intervallskala. Men hur är det med social önskvärdhet? Det är lätt att genomskåda en del uppgifter av den här typen. Betrakta denna:

*Jag blir nervös när jag ska tala inför en grupp*

Vem vill medge det i en anställningssituation? Det finns åtskilliga bevis för att många människor inte är helt uppriktiga när de svarar på ett test om resultatet kan bli viktigt för dem: att få ett jobb, komma in på en utbildning o dyl. Detta är ett kärnproblem i personlighetstestning, ofta kallat "faking". Hur ska man hantera den problematiken?

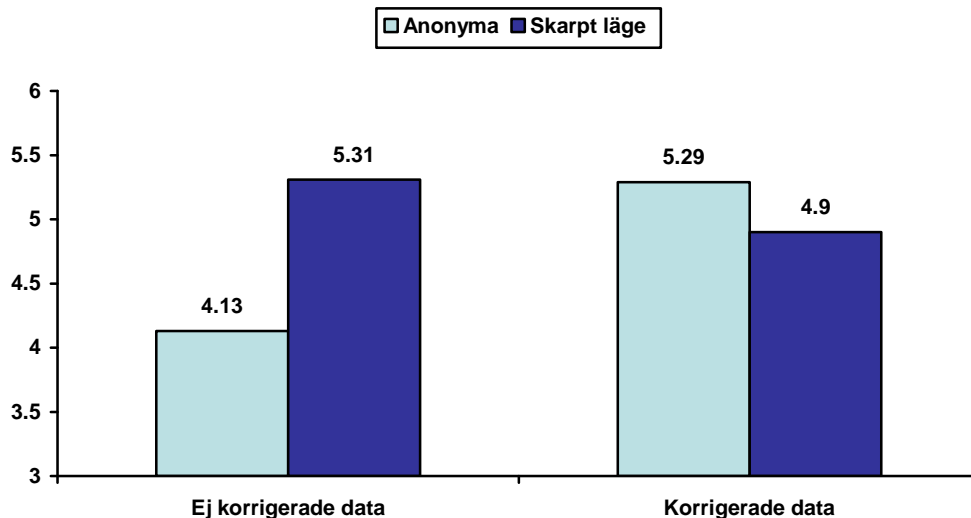
Det finns två sätt:

1. Gör en separat mätning av tendensen att svara socialt önskvärt, t ex med Crowne-Marlowe skalan (Crowne & Marlowe, 1960) och korriger sedan data utifrån resultat med sådana skalor
2. Gå över till ett format där de testade gör jämförande bedömningar, ett s.k. ipsativt format

Jag ska nu diskutera dessa två angreppssätt.

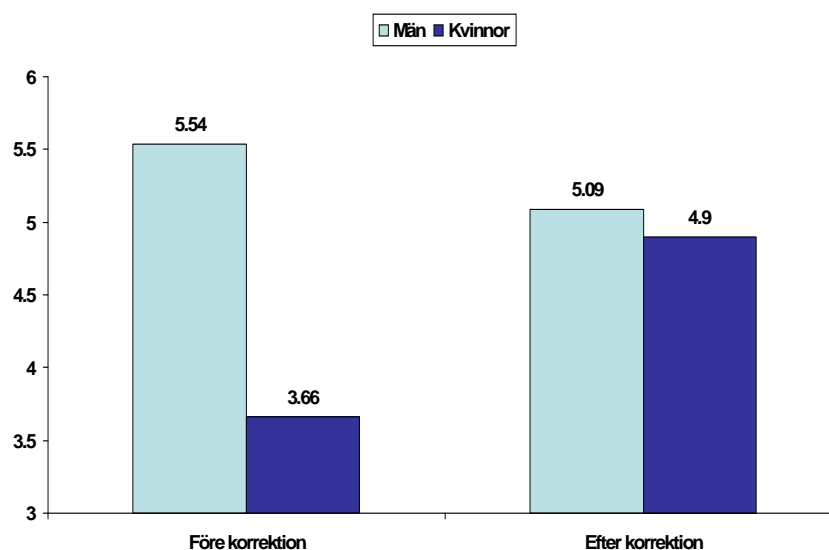
#### *Korrektion av normativa testdata som lösning av problemet*

Att mäta separat, med en eller flera skalor, den tendens de svarande har att vara taktiska i sina svar har visat sig fungera bra. Man kan korriger bort effekterna av skönmålningen, som ju kan ha stora effekter i enskilda fall. Ett exempel kan ges från en studie av sökande till Handelshögskolan, 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 1 visar resultatet av korrektion av personlighetsdimensionen emotionell stabilitet.



**Figur 1. 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 Figur 2 som ger könsskillnader före och efter korrektion i en grupp av personer som var kandidater till chefsjobb.



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

Figur 2 visar en mycket stor fördel för männen före korrektion för skönmålning. Efter korrektion försvann den nästan helt. Resultatet beror på 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 mindre än personer av svensk härkomst och missgynnas därför av test som inte korrigerar för skönmålning.

Forskningen om skönmålning svar på personlighetstest har i allmänhet använt sig av en experimentell uppläggning och har inte gett lika tydliga resultat (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 viss förvanskad bild av sig själv.

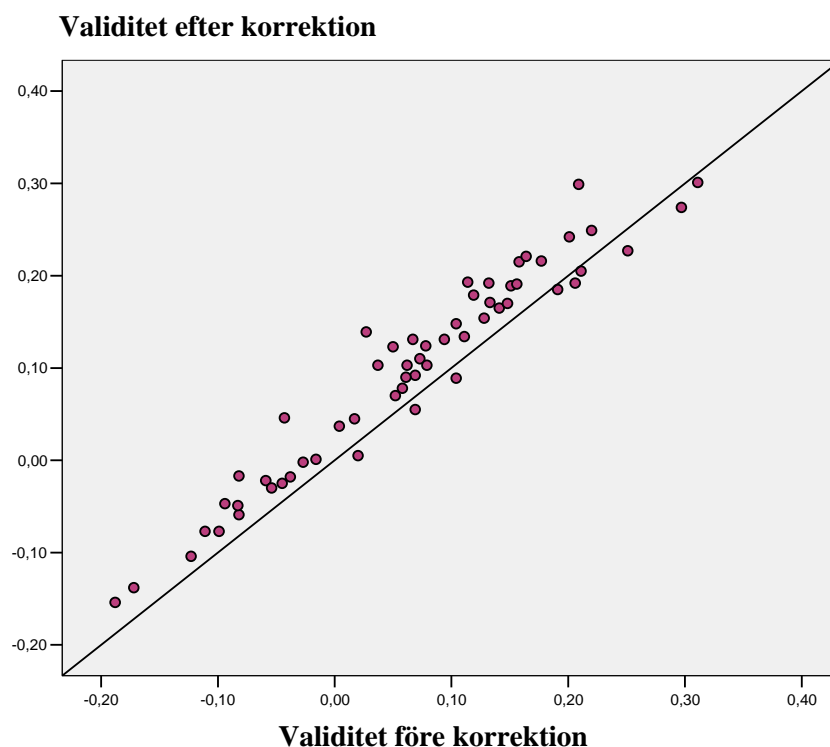
Den traditionella Crowne-Marloweskalen består av uppgifter som frågar om man brutit mot normer på sätt som är ganska vanliga. Förnekar man det så svarar man troligen taktiskt. Jag har konstruerat en egen skala utifrån samma idéer.

En person med kunskaper om psykologisk testning kan kanske genomskåda en traditionell skala för att mäta skönmålning. Jag har därför konstruerat en skala som troligen är omöjlig att genomskåda. Resultaten av olika sådana "lögnskalor" är likartade, enligt vad jag funnit, vid omfattande empiriska jämförelser mellan dem. Det gäller emellertid att göra en statistisk korrektion som fullt ut utnyttjar informationen i data. Det räcker inte med att tillämpa en "cut-off"-regel, där man avvisar de sökande som har en hög poäng, säg de 25 % högsta, på en "lögnskala". Det är inte effektivt nog. Det mesta av problemet kvarstår.

Det finns en del forskning som visat att korrektion av den typ jag förespråkar inte leder till högre validitet hos testerna, men inte heller till lägre (Ones & Viswesvaran, 1998). Spelar 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. Många av dessa är män eftersom kvinnor är mindre benägna att bluffa i dessa situationer.

Härtill kommer att aktuell forskning visat att korrektion faktiskt kan ge högre validitet, se Fig. 3 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 3. Validitet hos skalor i *UPP*-testet efter korrektion för skönmålning prickade mot motsvarande data före korrektion.**

Korrektion med hjälp av skalor för mätning av skönmålning 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. Man kan inte göra en global bedömning av om en person är trovärdig. Det är något som varierar beroende på vilken personlighetsegenskap man diskuterar.

Normativa format medför vissa problem utöver skönmålning. Det har hävdats att normativa svarsformat kan medföra vissa besvärliga felkällor: att personer varierar i tendens att hålla med om påståenden oavsett innehåll (jasägartendens) och vad gäller den intra-individuella variationen i svar (A. Brown & Bartram, 2009). Somliga utnyttjar skalan fullt ut, andra har en

tendens att använda bara några få svars kategorier. Dessa svarstendensers samband med testskalorna reduceras emellertid starkt efter korrektionen för skönmålning (Sjöberg, 2009).

### *Ipsativt format som försök att lösa problemet*

Det ipsativa formatet är den mest populära försöket att lösa problemet med skönmålning. Vanliga exempel på detta förfarande är 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 nämligen ett flertal problem med ipsativa testformat.

Ett första problem är att det är mycket 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 uppgiftsanalys. Det går inte heller att studera faktorstrukturen i en uppsättning ipsativa skalor.

När man använder ipsativa format ger man en uppgift som ibland är 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 (2002) jämförde de testades reaktioner på ipsativt och normativt format av OPQ. De skriver:

“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. Formatet har inga kända fördelar framför det vanliga normativa. Det tycks, till priset av lägre reliabilitet, längre testtid och negativa reaktioner hos de testade, ge samma information som vanliga Likertskalor. Rent teoretiskt kan man tänka sig att Likertformatet är sämre än ”forced choice” eftersom testpoängen vid användning av Likertskalor också avspeglar nivån på hur mycket eller litet man håller med om påståendena, inte bara deras rangordning som vid ”forced choice”. Det är en empirisk fråga om detta är en fördel eller nackdel.

Personlighetstester används ofta som beslutsunderlag, där personer ska jämföras med varandra. Men ipsativa testpoäng är *definitions*mässigt omöjliga att jämföra mellan personer. En sådan jämförelse är därför meningslös, men tyvärr verkar det som om en del testanvändare antingen inte förstått detta. Modellanalys (Anna 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 ett en 10-kampare kommer 1:a i 9 grenar men 2:a i höjdhopp. Ipsativt sett är han en dålig höjdhoppare!

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, Martin, & Hunt, 2002). Uppläggningsen i de citerade studierna var experimentell. Det var alltså ingen reell antagningstestning som de 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 taktisksvar. I ett mera aktuellt arbete fann man att ett ipsativt format inte löste problemet med skönmålning svar (Heggstad, Morrison, Reeve, & McCloy, 2006).

### *Slutsatser*

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. Så skriver t.ex. Meade (Meade, 2004):

“In sum, the standards required of tests used for employee 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.” (sid. 549).

Man bör alltså undvika test som använder med 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. Man uppnår inte den fördel man eftersträvade och man får en mängd andra allvarliga problem på köpet. Skönmålning kan effektivt hanteras vid användning av normativa format. Ipsativa format har visat sig vara en återvändsgränd i testmetodiken. Ju förr de överges desto bättre.

### **Referenser**

- 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.
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. [doi:10.1177/0013164410375112]. *Educational and Psychological Measurement*, 71(3), 460-502.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting and Clinical Psychology*, 24, 349-354.
- 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.
- Heggstad, 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. *Journal of Applied Psychology*, 91, 9-24.
- Mabon, H. (2004). *Arbetspsykologisk testning. Om urvalsmetoder i arbetslivet. Andra upplagan*. Stockholm: Psykologiförlaget.

- 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.
- 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.
- Sjöberg, L. (2009). *UPP-testet: Korrektion för skönmålning. (The UPP test: Correction for impression management)*. *Forskningsrapport 2009:3*. 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).

## Fullständig referenslista

- Arvonen, J. (2002). *Change, production and employees. An integrated model of leadership*. Stockholm: Department of Psychology, University of Stockholm.
- 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.
- 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. [doi:10.1177/0013164410375112]. *Educational and Psychological Measurement*, 71(3), 460-502.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting and Clinical Psychology*, 24, 349-354.
- 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.
- 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.
- 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. *Journal of Applied Psychology*, 91, 9-24.
- Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 74, 167-184.
- Mabon, H. (2004). *Arbetspsykologisk testning. Om urvalsmetoder i arbetslivet. Andra upplagan*. Stockholm: Psykologiförlaget.
- 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.
- 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.
- Sackett, P. R., & Lievens, F. (2008). Personnel selection. [doi:10.1146/annurev.psych.59.103006.093716]. *Annual Review of Psychology*, 59, 419-450.
- Sjöberg, L. (2009). *UPP-testet: Korrektion för skönmålning. (The UPP test: Correction for impression management)*. *Forskningsrapport 2009:3*. 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).