

Ny metod att avslöja skönsmålning vid personlighetstestning

Hur hanterar man att människor inte är uppriktiga vid testning och gärna skönsmålar? En vanlig metod för att komma tillrätta med problemet är att använda jämförande, eller ipsativa, svarsformat.

Men det finns bättre metoder, skriver professor Lennart Sjöberg, som konstruerat en egen skala som inte så lätt kan genomskådas.

Skönsmålning 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 nackdelar. Den medför besvärliga psykometriska problem, och upplevs negativt 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 skönsmålning. Det kan man emellertid göra till åtminstone 90 procent 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 skönsmålar mindre än andra grupper missgynnas de om man inte korrigerar.

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 (Likert-



Professor Lennart Sjöberg.

format). Formatet 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. Åtskillig forskning ger stöd för ett sådant sätt att arbeta (Sjöberg, 2006). 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 anställnings-

situation? 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 och dylikt. Detta är ett kärnproblem i personlighetstestning, ofta kallat "faking" (Griffith & Peterson, 2006). Hur ska man hantera den problematiken?

Det finns två sätt:

1. Gör en separat mätning av tendensen att svara socialt önskvärt, till exempel 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 så kallat 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 skönsmålningen, som ju kan ha stora effekter i enskilda fall.

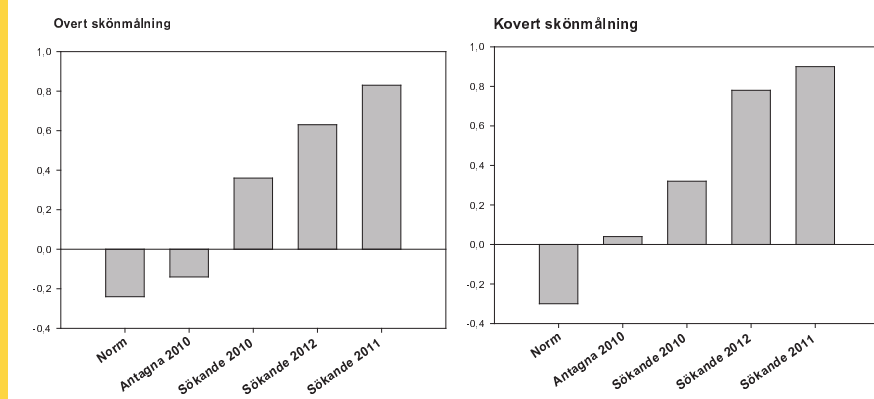
I vår forskning om sökande till officersutbildning vid Försvarshögskolan har vi haft tillfälle att studera olika grupper i varierande grav av "skarpt läge" (Sjöberg & Wolgers, in press). I normgruppen var läget helt oskarpt,

Skönmålning

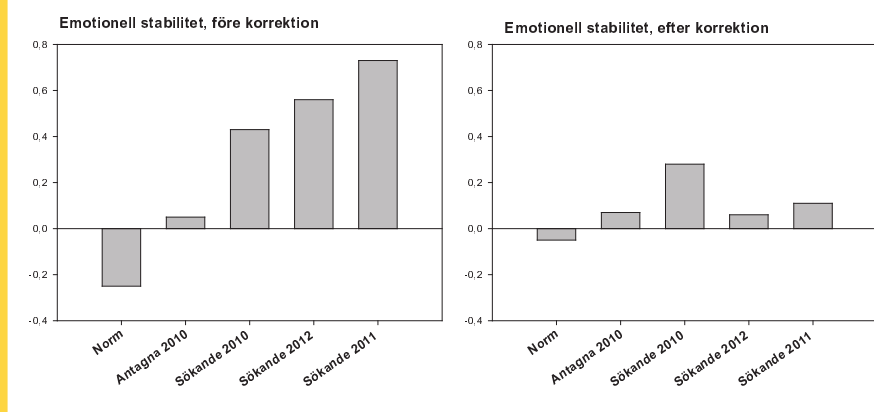
något liknande gäller redan antagna. Sökande 2011 tog testet i samband med ansökan och hade inget skäl att tro annat än att det skulle räknas in i beslutsunderlaget. Sökande 2010 och 2012 fick olika skarpt formulerade upplysningar om att testningen inte skulle tas med i beslutsunderlaget.

Två skalor för att mäta skönmålning användes: en overt och en kovert. Den overta skalan är en traditionell skala för mätning av social önskvärdhet av Crowne-Marlowetyp (Crowne & Marlowe, 1960). 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. Den koverta skalan kan inte på samma sätt genomskådas av en person som är inläst på testteori. Den består av testuppgifter som är av samma typ som vanliga personlighetsfrågor, men som är utvalda för att de har speciellt höga samband med den overta skalan. 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 procent högsta, på en "lögnskala". Det är inte effektivt nog. Det mesta av problemet kvarstår.

Fig. 1 visar att medelvärdena i dessa skalor var starkt relaterade till graden av skarpt läge: ju skarpare läge desto mera skönmålade de testade personerna. Trenden är densamma för overt och kovert skönmålning.



Figur 1. Medelvärden i skönmålning för fem olika grupper som testades i varierande grad av skarpt läge.



Figur 2. Medelvärden i emotionell stabilitet före och efter korrektion för skönmålning.

Det är intressant att notera de stora skillnaderna i skönmålning, eller social önskvärdhet, som följer graden av skarpt läge. En vanlig uppfattning om att sådana skalor mäter "personlighet" (Crowne & Marlowe, 1964, Uziel, 2010) är oförenlig med dessa resultat. Det som mäts är en felfaktor i testsvaren,

precis som avsetts.

För korrektion fanns stora skillnader i sådana variabler som emotionell stabilitet. De försvann i stort sett efter korrektion, se Fig. 2.

Ett annat 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. Det fanns en stor skillnad till förmån för de testade i skarpt läge, som efter korrektion försvann och till och med ändrades till en liten fördel för dem som testades helt anonymt.

Hur viktigt det kan vara med korrektion fann vi i en annan studie av könskillnader före och efter korrektion i en grupp av personer som var kandidater till chefsjobb.

Data visade på 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 (Sjöberg, 2010).

Forskningen om skönmålning svar på personlighetstest har i allmänhet använt sig av en experimentell uppställning 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.

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. Dessutom är den traditionella validiteten ett okänt mått när det gäller att avgöra om korrektion lett

till förbättringar av testet.

Aktuell forskning har visat att korrektion faktiskt kan ge högre validitet, något som vi funnit i data 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).

»Invandrare har visat sig skönmåla mindre än personer av svensk härkomst och missgynnas därför av test som inte korrigerar för skönmålning«

Vår 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 mycket 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 svarsalternativ. 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 po-

pulära 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 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 den testade 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. 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 *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 (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 exempelvis 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 nio 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 antagnings-testning 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 taktiska svar. I ett aktuellt arbete fann man att ett ipsativt format inte löste problemet med skönmålning svar (Heggstad et al, 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 (Converse et al, 2006). Så skriver till exempel 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 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. ☀

LENNART SJÖBERG

Professor em i psykologi vid Handelshögskolan i stockholm

Fullständig referenslista kan beställas från redaktionen, e-post: tidningen@psykologforbundet.se

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.

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.

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.

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.

Sjöberg, L., Bergman, D., Lornudd, C., & Sandahl, C. (2011). *Sambandet mellan ett personlighets test och 360 graders bedömningar av chefer i hälso- och sjukvården*. Stockholm: Karolinska Institutet, Institutionen för lärande, informatik, management och etik (LIME).

1 månad kvar ...

... för dig som inte är klar med din specialistutbildning enligt den gamla ordningen.

☀ Den gamla ordningen gäller dig som är antagen före oktober 2007. Slutdatum är 31/12/2012.

Det gäller för medlemmar som har kompletteringsbesked sedan mitten av 1990-talet, eller som startade sin specialistutbildning från grunden före oktober 2007.

Vad innebär det att "bli klar"?

– Att vara färdig med alla moment som ingår; professionkurs, specialistkollegium, sex specialistkurser samt det vetenskapliga arbetet,

säger Håkan Nyman, specialistrådets ordförande.

Har du frågor, vänd dig till Specialistrådet (se *Psykologtidningen* näst sista sidan) eller till professionsansvarig Maria Lindhe, tel 08-567 06 477.