



PSYKOLOGISK METOD AB



UPP-testet: tredje generationens personlighetstest

Lennart Sjöberg
Rapport 2010:1

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

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

Skrifter utgivna av Psykologisk Metod AB

Sjöberg, L. (2008). Bortom Big Five: Konstruktion och validering av ett personlighetstest. Rapport 2008:1.

Sjöberg, L. (2008). *Big Five Plus*. Manual.

Sjöberg, L., & Möller, K. (2009). Sociala arbetsfunktioner och personlighet. Rapport 2009:1.

Sjöberg, L. (2009). *UPP*-testet: Kriterierelaterad validitet. Rapport 2009:2.

Sjöberg, L. (2009). *UPP*-testet: Korrektion för skönmålning. Rapport 2009:3.

Sjöberg, L. (2010). *UPP*-testet: Tredje generationens personlighetstest. Rapport 2010: 1.

Sammanfattning

Utvecklingen av personlighetstestning inom arbetslivet har genomgått tre faser. Den första generationens test, exempel Cattells 16 pf eller OPQ, utmärktes av mycket komplicerade system för beskrivning av personligheten. Dessa system förenklades delvis i den följande generationen av test, som byggde på femfaktormodellen, men den modellen var enkel bara på den övergripande nivån. Under de fem huvudfaktorerna fanns ett stort antal underordnade faktorer, i allmänhet 30-40 till antalet. Inga test i första och andra generationen kunde effektivt hantera problemet med skönmålande svar, inte heller tog man hänsyn till effekterna av stämningssläge på testsvaren. Dessa och ett antal andra problem löstes till stor del i UPP-testet, som därför föreslås representera den tredje generationen av personlighetstest.

Inledning

Syftet med denna uppsats är att presentera nya ansatser till personlighetstestning i arbetslivet. Dessa har förverkligats i *UPP*-testet, som sätts i relation till tidigare utvecklade test. För en historisk översikt av arbetspsykologisk testning, se (Sjöberg, 2004). Tre testgenerationer kan urskiljas, med början i 1920-talet.

Första generationen: 1920-1990

Självrapporttest för personlighet går tillbaka till första världskriget. Fram till 60-talet dominerade i den amerikanska traditionen faktortest som Cattells 16pf – 16 faktorer - eller Eysencks betydligt enklare MPI, som bara hade 3 faktorer. Den amerikanska traditionen kom emellertid att dominera. Det blev därför tämligen komplicerade test som var vanligast. Men en forskare som Cattell, skapare av 16pf, byggde sitt test på faktoranalyser av små datamaterial; resultaten har aldrig kunnat replikeras (Cooper, 2002). Att den första generationens personlighetstest utmynnade i mycket komplicerade teststrukturer berodde alltså troligen delvis – kanske helt - på att man inte hade tillgång till effektiv statistisk och psykometrisk metodik.

Vissa äldre test som är mycket ofta använda i Sverige, PPA och OPQ, bygger på äldre och i fallet med OPQ mycket komplicerade modeller. OPQ32r mäter 32 personlighetsvariabler med ganska få uppgifter, totalt 104 block med 3 påståenden i varje block, där man ska markera i varje block vilket som stämmer bäst och vilket som stämmer sämst. Tillförlitligheten för 32 skalor så få uppgifter riskerar att bli låg, och faktorstrukturen har aldrig kunnat replikeras. OPQ bygger på en common-sense analys som gjordes för snart 30 år sedan. En meta-analys av brittisk forskning om OPQ (Robertson & Kinder, 1993) visade att testet hade en prediktiv validitet omkring 0.20, vilket är betydligt sämre än de flesta andra test som brukar ligga på 0.3, enligt Mischels klassiska studie (Mischel, 1968). I den senaste utvecklingen av OPQ har det visat sig att testet har en femfaktorstruktur (Brown & Bartram, 2009), varom mera i nästa avsnitt.

Den praktiska tillämpningen av testanvändningen förändras bara ytterst långsamt. Det mest använda testet i Sverige är PPA, en adjektivlista med rötter i 20-talet (Marston, 1989/1928):

70 000 testningar per år! Den mäter möjligen några viktiga dimensioner men långtifrån alla av betydelse i arbetspsykologiska tillämpningar. Se Mabons utmärkta diskussion av denna typ av test (Mabon, 2004). Ett annat äldre test är MBTI som bygger på Jungs personlighetsteori från början av 1920-talet. Se en kritisk diskussion som jag publicerat på annat ställe (Sjöberg, 2005).

De flesta av den första generationens test hade alltså nackdelen att de gav en ytterst komplex bild av personligheten. De ställde stora krav på erfarenhet och intuitiv förmåga hos de psykologer som använde dem. Efter en lång tids användning och erfarenhet av att testa många typer av personer kunde de i händerna på skickliga psykologer vara av värde, men valideringsforskning gav oftast som resultat att validiteterna var låga, 0,3 var en siffra som tycktes beskriva var taket låg. OPQ tycks ligga betydligt lägre, också enligt leverantörens egen senaste analys (Brown & Bartram, 2009).

0,3 är en siffra som Mischel gjorde en stor affär av i sin klassiska bok från 1968 (Mischel, 1968), men redan 1921 fann bröderna Allport samband på samma svaga nivå (Allport & Allport, 1921). De trodde då att det skulle gå att få mycket starkare resultat med bättre test. Det har senare visat sig att de hade fel.

Ett av grundproblemen med de första generationernas personlighetstest var att man var upptagen av prediktion. Detta är i och för sig en hederlig ambition, men den är orealistisk. Även om vi har mycket effektiva och relevanta mått på personlighetsdimensioner finns ingen garanti för att vi kan göra prognoser av hur en person kommer att lyckas i en viss miljö eller med en viss uppgift. Det är ju många aspekter utom hans eller hennes personlighet som spelar in. Vi vet inte ens hur mycket av beteende eller framgång som i princip kan prediceras utifrån individfaktorer – frågan ställs mycket sällan på det sättet.

Andra generationen: 1990 - 2009

Omkring 1990 slog begreppet ”Big Five” igenom på bred front (McCrae & Costa, 1987). Det tycktes vara mycket attraktivt att det skulle räcka med bara fem faktorer för att beskriva personligheten. Amerikanska test som NEO-PI-R (Costa & McCrae, 2003) och HPI (Hogan, 1992) bygger på denna modell. Den internationella databasen IPIP har publicerat testuppgifter som fritt får översättas och användas¹. Den ger mått på de övergripande fem stora personlighetsdimensionerna. Dessa är:

- Utåtvändhet
- Emotionell stabilitet
- Noggrannhet
- Öppenhet
- Vänlighet

Big Five-testen är andra generationens personlighetstest. De är enklare än första generationens test i den meningen att de fokuserar på ganska få övergripande dimensioner. Men komplexiteten finns ändå kvar i form av underordnade skalor, kallade fasetter i NEO-PI-R; det handlar om 30-40 underordnade skalor och komplexiteten är alltså betydligt större än i

¹ <http://ipip.ori.org/ipip/>

sådana test av första generationen som 16 pf och OPQ, för att inte tala om de strukturellt enklare MBTI och PPA.

Andra generationens Big Five-test har haft ett starkt genomslag, speciellt i den internationella forskningen. Nya test bör utvärderas mot Big Five för att se om de har något nytt att ge, hävdas det ofta. Omfattande studier har nu gjorts av Big Five-dimensionernas praktiska värde i arbetslivet. Resultatet var emellertid nedslående (Morgeson, et al., 2007a, 2007b). Big Five-dimensionerna tycktes inte alls ha medfört någon bättre prognosförmåga.

Tredje generationen: 2009 -

Första och andra generationernas personlighetstest stod inför problem som de inte lyckades lösa och som först i en ny testgeneration kan hanteras effektivt. Jag ska nu behandla dessa problem och börjar med skönmålning. Mina flesta exempel kommer från det nya *UPP*-testet. *UPP* står för *Understanding Personal Potential*.

Beskrivningsnivå

Andra generationens test, Big Five, har skapat en paradoxal svårighet: de är dels för generella, dels för detaljerade. På den övergripande Big Five-nivån ger de testvärden som har ytterst svaga samband med relevanta kriterier, men på den detaljerade nivån med underordnade skalor, t ex "fasetter" i NEO-PI-R, ger de för många skalor som användaren inte kan hantera på ett effektivt sätt. Trettio två eller 40 skalor är för mycket att hantera kognitivt och tveksamhet uppstår om alla dessa skalor verkligen kan mätas reliabelt med en testtid på 30-40 minuter.

UPP-testet mäter, förutom Big Five, ett begränsat antal dimensioner som är utvalda för sin speciella relevans i arbetslivet, och för att de är lätta att förstå:

Samarbetsvilja
Kreativitet
Positiv grundattityd
Uthållighet
Självförtroende
Social förmåga
Emotionell intelligens
Perfektionism

I relation till proxy-kriterierna svarade dessa fokuserade variabler för ca 8 gånger större andel av variansen än Big Five. Multipla korrelationer för vart och ett av de sex kriterierna mot FFM-modellen Big Five och mot *UPP*-testet framgår av Tabell 1. I *UPP*-testet ingick också de 8 tillkommande variablerna – uthållighet, samarbetsvilja etc.

Tabell 1. Validering mot 6 kriterievariabler, förklarad varians (kvadrerade multipla korrelationer).

Kriterievariabel	Big Five	UPP-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

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

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

Skönmålning

Varken första eller andra generationen av personlighetstest lyckades bemästra problemet med skönmålning. Skönmålning är ju ett uppenbart problem för alla självrapporttest. Det finns tre sätt att hantera det:

Förneka att det existerar. Omfattande forskning visar att skönmålning är vanligt förekommande och att den kan ha enormt stora effekter i enskilda fall.

Arbeta med ipsativa testuppgifter, där den testade ska välja mellan svarsalternativ som är matchade i social önskvärdhet. Problemen är många med denna vanliga ansats. Matchningen i social önskvärdhet är givetvis oprecis och gäller inte genomgående för alla som tar testet. Forskning visar att skönmålning ändå slår igenom. 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; Martin, Bowen, & 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). Dessutom är den psykometriska analysen mycket svår för sådana data. Vidare: de testade tycker inte om ipsativa test – de upplever att de måste "jämföra päron och äpplen" – och det tar mera tid att genomföra testet (Harland, 2003). Testresultaten kan sedan inte jämföras *mellan* individer, bara *inom*. Man kan alltså inte säga "Kalle är mera kreativ än Pelle", bara "Kalle är mera

kreativ än han är utåtvänd”. Test av den typen kan i princip inte användas vid urval (men det sker ändå). 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. 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 arbetar 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.

Konstruera och använd någon form av ”lögnskala”, eller skala på tendensen att svara på ett socialt önskvärt sätt, som ger ett särskilt mått på tendensen att skönmåla. Denna skala kan användas för att i en urvalssituation sortera bort dem som ha högt värde, men den strategin medför inte att effekten av skönmålning försvinner, den bara mildras något (Sjöberg, 2008). Lögnskalan används globalt för en hel testning, medan forskningen visat att skönmålning är av ytterst varierande betydelse beroende på vilken testvariabel som studeras. Tabell 2 ger ett exempel från en studie av ett stort datamaterial över jobbsökande som tagit ett test i samband med ansökan om jobb (Sjöberg, 2009a). Korrelationerna mellan personlighetsvariabler och mått på skönmålning är som synes tämligen varierande.

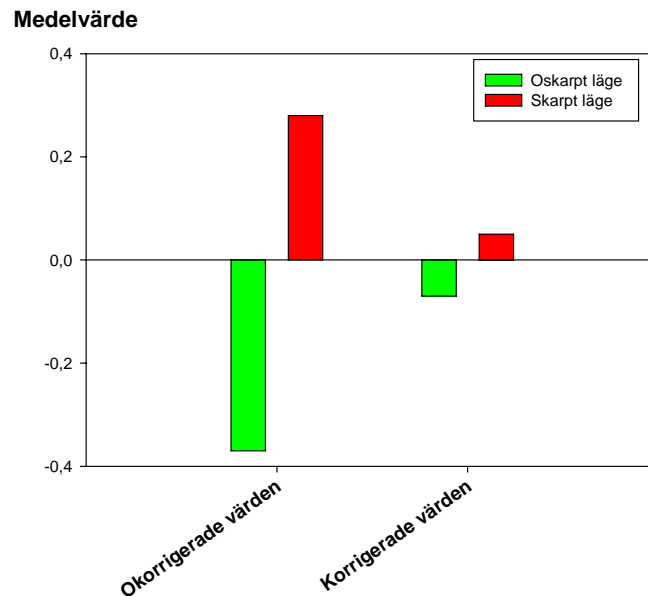
Tabell 2. Variansandel som förklaras av skönmålning (N=2202)

Testvariabel	Andel av skönmålning förklarad varians
Extraversion	0.213
Uthållighet	0.398
Samarbetsvilja	0.419
Positiv grundattityd	0.381
Kreativitet	0.089

I tredje generationens personlighetstest används avancerad psykometrisk och statistisk metodik för att korrigera för skönmålning i varje enskild testvariabel – allt eftersom data visar att korrektion behövs och hur mycket. Denna metod bygger på att man mäter tendensen att skönmåla med en eller flera skalor som är insprängda bland övriga testuppgifter.

Regressionsanalyser görs av varje testskala mot måtten på skönmålning och residualerna beräknas genom differenser mellan erhållna och predicerade värden i testskalorna. Dessa residualer är definitionsmässigt nollkorrelerade med skönmålning som den definierats med

hjälp av de särskilda måtten. I min forskning har jag tillämpat denna metodik i olika sammanhang: experimentella och fältstudier. Ca 90-95 % av effekterna av skönmålning försvinner. Man kan alltså inte ”lura” testet att tro att man har egenskaper som man faktiskt inte har. Se Fig. 2 för resultatet av korrektion för skönmålning i de jobbsökardata som också användes i Tabell 2.



Figur 2. Genomsnittliga värden i personlighetsvariablerna före och efter korrektion för skönmålning, skarpt och oskarpt läge.

Validering mot proxy-kriterier

Var och en som sysslat med att konstruera och validera test har stött på problemet med att få in data på kriterier av typ arbetsprestation. Företag är sällan positiva till att ta fram och lämna ut sådan information om sina anställda. STP² kräver dessutom att man måste ha en grupp om minst 100 personer för en validering mot arbetsprestation vilket gör det nästan omöjligt att validera ett test på det sättet.

Nu finns det i ett tredje generationens test en mycket enklare lösning, nämligen att använda *proxy-kriterier*. Det finns nämligen ett antal attityd- och motivationsdimensioner av typ arbetsvilja och arbetstillfredsställelse som vi vet är relaterade till arbetsresultat. Dessa dimensioner kan användas som ställföreträdande kriterier och de kan mätas i samma session som testningen och korrigeras för skönmålning på samma sätt som de vanliga testvariablerna. *UPP*-testet mäter:

- Arbetsvilja
- Arbetstillfredsställelse
- Arbetsintresse
- Resultatorientering

² Stiftelsen för Tillämpad Psykologi, som gör granskningar av psykologiska test.

- Förändringsvilja
- Balans arbete/övrigt liv

Dessa dimensioner används för att validera testet men de har också i många sammanhang ett egenintresse och många av testets användare ser i dem ett unikt tillskott som saknas i personlighetstest av första och andra generationen. Skulle de av någon anledning inte vara av intresse i en speciell tillämpning kan de givetvis uteslutas, vilket görs i en av *UPP*-testets 7 moduler (mera om modulerna nedan).

Användningen av proxy-kriterier utesluter inte att vi också använder mera konventionella mått på arbetsprestation för utvärdering av testet. I en studie av chefskarriär fann vi en korrelation på 0,54 mellan *UPP*-testet och chefsansvar (Sjöberg, 2009b).

Stämningläge

Det är rimligt att tro att de svar man ger på ett självrapporttest påverkas av det emotionella tillstånd man befinner sig i, av stämningläge eller ”mood”. Det har länge varit känt hur stämningläge kan mätas och att tre enkla dimensioner räcker för att beskriva det: hedonisk ton (glad-ledsen), spänning (lugn – nervös) och aktivering (pigg - trött), se (Sjöberg, Svensson, & Persson, 1979). Få tycks emellertid ha undersökt närmare hur stämningläget påverkar svaren i ett personlighetstest. Kanske beror detta på att man har antagit att sådana svar är tämligen konstanta över tid. Men detta är en halvsanning; över en tidsperiod på ca 6 veckor är korrelationerna ca 0,7 mellan två testningar med samma test (Sjöberg, 2008). Detta resultat, som är typiskt för självrapporttest av personlighet, betyder att det finns en viss stabilitet men också att det förekommer variationer och att dessa kan vara stora i enskilda fall.

Det är därför viktigt att undersöka vilket känslotillstånd den testade personen befinner sig i när testet tas. Om detta är alltför avvikande finns anledning att vara försiktig med tolkningen av resultaten, eller att göra om testet. Se Tabell 3 och 4 samt Fig. 3 för sambandet mellan humöret och de vanliga fem stora personlighetsdimensionerna. Dessa data kommer från en starkt engagerande testsituation vid antagningar till Handelshögskolan i Stockholm. Med ”humör” menas ett sammanslaget mått som kombinerar alla tre dimensionerna som mäter stämningläge.

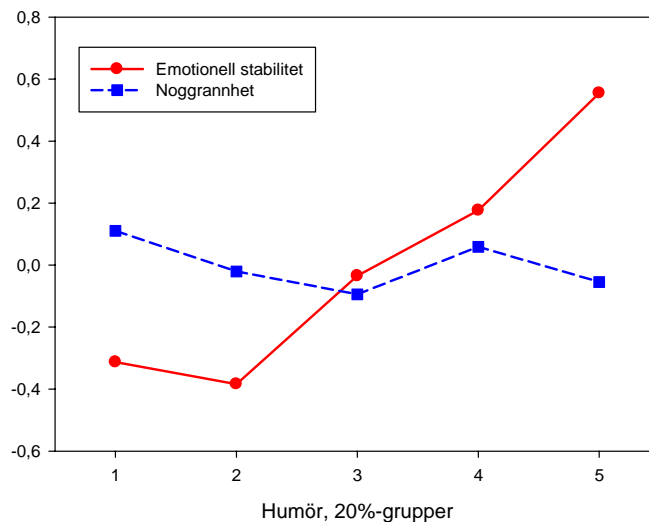
Tabell 3. Korrelationer mellan stämningläge och Big Five-dimensionerna, ej korrigerade för skönmålning.

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

Tabell 4. Korrelationer mellan stämningläge och Big Five-dimensionerna. Alla värden korrigerade för skönmålning.

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

Medelvärde i personlighetsskala, standardiserad



Figur 2. Emotionell stabilitet och noggrannhet som funktioner av stämningläge vid testningen.

Tabellerna 3 och 4 visar på starka samband mellan stämningläge och tre av de fem personlighetsvariablerna, *särskilt om de senare inte korrigerats för skönmålning*.

Fig. 2 visar stora skillnader i emotionell stabilitet mellan dem som låg i gruppen bestående av de 10 % som hade ”sämst humör” vid testningens början, och övriga. Med ”sämst humör” menas att de var mera nervösa, nedstämda och trötta än övriga.

Första och andra generationens test har inte beaktat denna problematik, som är uppenbart viktig i vissa fall. *UPP*-testet inkorporerar en mätning av stämningläge i början av testningen. Resultatet sätts i relation till normdata och tolkas som en aspekt på datakvalitet. Det finns emellertid flera viktiga aspekter på datakvalitet.

Datakvalitet

Om man kan dra slutsatser från resultatet på ett personlighetstest beror på datas kvalitet. Det är ett stort problem vid testning att kvaliteten varierar från person till person. Somliga testade har haft alltför lågt engagemang i testningen eller missförstått uppgifterna på ett eller annat sätt. Då blir resultaten osäkra eller kanske helt oanvändbara. Det är därför enormt viktigt att mäta testmotivationen. I vår forskning har vi lyckats ta fram metodik för sådan mätning. Det är en bred ansats som löser problemet att mäta datakvalitet på delvis nya sätt.

Hur mäter man då kvalitet? Det finns ett flertal aspekter att ta hänsyn till:

- Skönmålning, dvs. taktiska svar som den testade gett för att ge en positiv bild av sig själv. I *UPP*-testet finns två skalor för att mäta skönmålning.
- Differentiering, dvs spridning av svaren; motsatsen är att den testade personen har haft en tendens att svara med samma alternativ på alla testuppgifter.
- Ja-sägartendens, dvs tendensen att instämma i testuppgifter oavsett deras innehåll.
- Slumpmässiga svar, eller svar som tyder på att den testade uppfattat testuppgifterna på annat än det avsedda sättet.
- Antal ej besvarade frågor.

Alla dessa kvalitetsdimensioner mäter vi reliabelt i *UPP*-testet. Kvalitetsdimensionerna har en tendens att samvariera. De vägs samman till en övergripande bedömning av testdatas kvalitet för varje testad person. Detta förfarande ger ett mått på testmotivation, som är högt om den testade personen noggrant besvarat alla eller nästan alla testuppgifter på ett sätt som tyder på att han eller hon förstått uppgifterna som avsett.

Information om testdatas kvalitet integreras i testrapporten för varje person. Om värdet på kvaliteten är lågt, t ex 3 eller mindre på den 9-gradiga stanineskalan som vi använder, finns anledning att behandla testresultaten mycket försiktigt eller inte alls ta hänsyn till dem, kanske att göra om testningen efter utredning av skälen till den låga kvaliteten.

Tolkningen av kvalitetsmått underlättas av att vi har omfattande normdata, ca 1000 testade personer; datas kvalitet och testmotivationen har beräknats för var och en.

Hos vissa andra test förekommer några av de mått vi använder, oftast i enklare form. Det förekommer också att man mäter "ovanliga svar" som ett mått på kvalitet. Men ovanliga svar behöver inte betyda att testningen är av låg kvalitet. En person med ett extremvärde i en testvariabel *måste* ge ovanliga svar för att uttrycka detta värde.

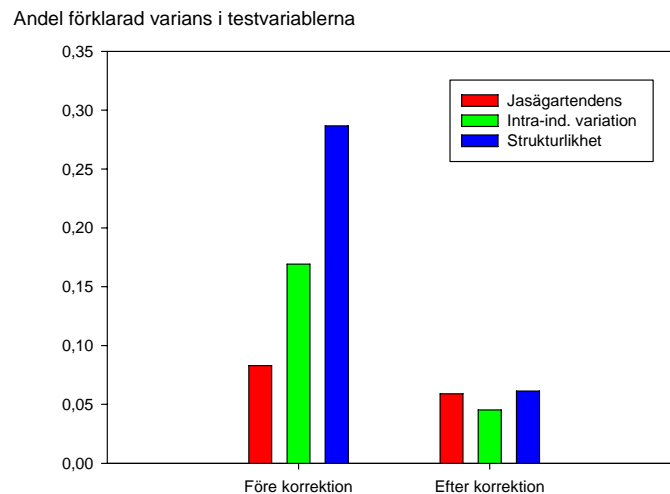
Som nämndes i inledningen har det hävdats att normativa svarsformat kan medföra vissa besvärliga felkällor: skönmålning, samt att personer varierar i tendens att hålla med om påståenden oavsett innehåll (jasägartendens) och vad gäller variationen i svar. Somliga utnyttjar skalan fullt ut, andra har en tendens att använda bara några få svars kategorier.

Det är sant att ipsativa format undviker problemen med ja-sägande och intra-individuell variation. Men hur är det med ett normativt format *med data korrigerade för skönmålning*? Frågan ska här analyseras med hjälp av data från normgruppen (N=953).

Korrelationer beräknades mellan vart och ett av de tre måtten på datakvalitet och alla testvariabler (19 stycken). Kvalitetsmåten var:

- Jasägartendens
- Intra-individuell variation
- Strukturlikhet

Alla korrelationer kvadrerades för att få ett mått på hur stor del av variansen i testvariablerna som förklarades av kvalitetsmåten. De genomsnittliga kvadrerade korrelationerna illustreras i Fig. 3.



Figur 3. Genomsnittlig andel av variansen i testvariablerna som förklaras av kvalitetsmåten jasägartendens, intra-individuell variation och strukturlikhet.

Figuren visar en mycket tydlig skillnad mellan data före och efter korrektion för skönmålning. Okorrigerade data kunde till ganska betydande del förklaras utifrån kvalitetsmått. Så kunde t ex ett högt värde i testvariablerna bero på att en person tenderat att välja olika svars-kategorier och inte grupperat sina svar till några få kategorier. Det är just den sortens resultat som anhängare av ipsativa testformat brukar anföra som kritik av normativa format. Men data efter korrektion för skönmålning förklaras bara till ca 6 % av kvalitetsmåten. Detta är en reduktion med ca 2/3 av effekten av kvalitetsvariationen.

Normativt vs ipsativt

De ipsativa testen har åtskilliga nackdelar, som diskuterats ovan. Men ipsativ utvärdering har en klar fördel, som inkorporerats i *UPP*-testet trots att detta är i princip normativt. Man kan nämligen ta fram ipsativ information ur ett normativt test men inte tvärtom. Det är just detta som görs i *UPP*, vars rapporter innehåller ett avsnitt som ger jämförelser inom individen helt bort sett från hur han eller hon ligger till i förhållande till normdata.

På detta sätt tar vi fram information om *relativt sett* starka sidor och utvecklingsmöjligheter. Det är data som kan vara lättare att visa i t ex grupparbete än de testvärden som grundas på jämförelser med en normdatabas.

Aggregerade variabler: Exemplet jagstyrka

Uthållighet, samarbetsvilja, emotionell stabilitet etc. bildar ett väl sammanhållet kluster av variabler. Vad är det som är gemensamt hos alla dessa olika personlighetsvariabler? Svaret är jagstyrka. Det är en term ur den kliniska psykologin, och den används sällan i arbetspsykologiska sammanhang. Mycket svag jagstyrka är en klinisk problematik, men även i en normalpopulation varierar jagstyrkan. Det är en term som lämpar sig väl för att beskriva en sammanfattande bedömning som baseras på ett stort antal delvariabler. Jagstyrka är förmågan att se sig själv och situationen på ett realistiskt sätt och att behärska kortsiktiga impulser.

UPP arbetar med flera sammansatta variabler utöver jagstyrka, nämligen stresskänslighet och chefspotential. Det är en fördel för testanvändaren att på detta sätt få ett mindre antal viktiga, övergripande dimensioner utvärderade.

Utvecklingen av personlighetstest av första och andra generationerna gick mot fragmenterade och komplicerade personbeskrivningar. Användarnas kommentar var ofta att de inte visste hur de skulle tolka dem.

Moduler

Ett tredje generationens test bör vara maximalt flexibelt och tillåta många olika typer av användning. Hela *UPP*-testet är ofta det som efterfrågas vid urval och rådgivning, men delar av testet kan vara mycket användbara i vissa sammanhang. Följande moduler finns:

- A. Hela testet
- B. Hela personlighetstestet minus måtten på arbetsmotivation och övriga proxykriterier
- C. Enbart mätning av social förmåga och emotionell intelligens
- D. Enbart proxykriterierna
- E. Screeningtest. En starkt förkortad del av ett urval av personlighetsvariablerna
- F. Förkortat test som enbart mäter jagstyrka

Narrativa rapporter – artificiell intelligens

Den första generationen av personlighetstest gav resultat som skulle tolkas av psykologen – en profil i de olika testvariablerna. Sådana tolkningar tenderar att bli subjektiva. Olika psykologer kommer till olika resultat med ett och samma underlag (Dawes, 1994). Senare, när datorerna kom in i bilden, skrevs program som skapade verbala, narrativa, rapporter med grundval i testprofilerna. *UPP*-testet skapar en narrativ rapport som bygger på följande principer:

Varje testvariabel delas i tre (i vissa fall fyra) nivåer avseende stanine-poäng. Testvariablerna grupperas i några övergripande teman. Inom varje tema återges en text som är skriven för just den nivå på testet som uppnåtts av en testad person. Texterna bygger på personlighetspsykologisk forskning. Sammantaget ger den narrativa rapporten en känsla av att man ser en verklig person framför sig och att man förstår honom/henne bättre än om man bara får en kvantitativ profil.

Attityd till testet: ” Face validity”

Hur den testade upplever testet och testningen är viktigt att veta vid tolkningen av resultaten och kan vara en värdefull ingång till återkopplingsintervjun. Av det skälet har frågor om attityden till testet medtagits i *UPP*, också det en unik fördel för detta test. Det är intressant att närmare studera hur de testade har bedömt *UPP*.

På frågan om testningen var meningsfull svarade 89,6 % ja, 8,3 % var tveksamma och 2,1 % svarade nej. Var testet bättre eller sämre än andra personlighetstest? Av dem som hade erfarenhet av andra test svarade 51% att *UPP* var bättre, 49 % att *UPP* var varken bättre eller sämre. Ingen svarade att *UPP* var sämre.

Slutsatser

Viktiga framsteg har gjorts inom utvecklingen av personlighetstest:

- Utvecklingen mot alltmera komplicerade personlighetsbeskrivningar bryts dels genom användning av ett mycket begränsat antal grundläggande dimensioner, dels genom introduktionen av ett fåtal aggregerade dimensioner av typ jagstyrka
- Problemet med skönmålande svar vid ett normativt svarsformat har till stor del lösts
- Tydliga effekter har dokumenterats av stämningssläge vid testning; mätningar av stämningssläge ingår därför som en del av testet
- Både normativa och ipsativa tolkningar av testresultaten är inkluderade i testrapporten
- Valideringsarbete har i hög grad underlättats tack vare ansatsen med proxy-kriterier
- Omfattande analys görs av datakvaliteten i varje enskild testning
- Testet kan ges i sin helhet eller i ett av flera modulformat, beroende på syftet med testningen
- Testet har ett inbyggt datorbaserat system för att skapa narrativa rapporter
- *UPP* innehåller frågor om hur testningen och testet upplevdes och har enligt resultaten god ”face validity”

UPP-testet inkorporerar dessa funktioner.

Referenser

- Allport, F. H., & Allport, G. W. (1921). Personality traits: Their classification and measurement. *Journal of Abnormal and Social Psychology*, 16, 6-40.
- 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, 240-259.
- Brown, A., & Bartram, D. (2009). *Development and psychometric properties of the OPQ32r. Supplement to the OPQ 32 technical manual*: SHL.
- Cooper, C. (2002). *Individual differences. 2nd edition*. London: Hodder Education.
- Costa, P. T., & McCrae, R. R. (2003). *NEO PI-R. Manual, svensk version av Hans Bergman*. Stockholm: Psykologiförlaget.
- Dawes, R. M. (1994). *House of cards. Psychology and psychotherapy built on myth*. New York: The Free Press.
- 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.
- 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.
- Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 74, 167-184.
- Hogan, R. (1992). Hogan Personality Inventory. *Psychological Test Bulletin*, 5(2), 130-136.
- Mabon, H. (2004). *Arbetspsykologisk testning. Om urvalsmetoder i arbetslivet. Andra upplagan*. Stockholm: Psykologiförlaget.
- Marston, W. M. (1989/1928). *Emotions of normal people*. Ormskirk, Lancs.: Thomas Lyster.
- Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences*, 32, 247-256.
- McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90.
- 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.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, 60(4), 1029-1049.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60(3), 683-729.
- Robertson, I. T., & Kinder, A. (1993). Personality and job competences: The criterion-related validity of some personality variables. *Journal of Occupational & Organizational Psychology*, 66, 225-244.

- Sjöberg, L. (2004). Personlighetstest i arbetslivet: historik och aktuell forskning. (Personality tests in industry: History and current research). In G. Sevón & L. Sjöberg (Eds.), *Emotioner och värderingar i näringslivet* (pp. 171-229). Stockholm: Ekonomiska Forskningsinstitutet - EFI.
- Sjöberg, L. (2005). En kritisk diskussion av Myers-Briggs testet. (A critical discussion of the Myers-Briggs test). *Organisational Theory & Practice. Scandinavian Journal of Organisational Psychology*, 15(1), 21-28.
- 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. (2009a). *UPP-testet: Korrektion för skönmålning*. Stockholm: Psykologisk Metod AB.
- Sjöberg, L. (2009b). *UPP-testet: Kriterierelaterad validitet*. Stockholm: Psykologisk Metod AB.
- Sjöberg, L., Svensson, E., & Persson, L.-O. (1979). The measurement of mood. *Scandinavian Journal of Psychology*, 20(1), 1-18.