

Måling av personlighet

Hilmar Nordvik og Trude Stokkan

Måling av personlighet

To metoder for å måle de samme personlighetstrekkene sammenliknes i denne studien. Den ene baserer seg på testskårer fra NEO PI-R, konstruert for å måle femfaktormodellen for personlighet. Den andre går ut på å be personer om selv å angi skårer som gjenspeiler deres personlighet.

Innledning

I moderne personlighetsforskning og personlighetsdiagnostisering er selvrapport en dominerende fremgangsmåte. Metoden forutsetter at personer selv kjenner sin personlighet slik at de kan informere om den gjennom sin rapport. Personen er den eneste som kjenner sin personlighet fra innsiden, og den eneste som alltid er til stede når han eller hun ter seg og viser seg fram.

Vi kan skille mellom to strategier i selvrapportering: *Adjektivmetoden* dreier seg om å vurdere seg selv direkte på ord som beskriver personlighetstrekk. Et eksempel er: «På en skala fra 1 til 5, hvor 'selskkelig' mener du at du er?» *Setningsmetoden* innebærer å vurdere seg selv i forhold til utsagn om tilbøyeligheter, for eksempel: «På en skala fra 1 til 5, hvor enig er du i utsagnet: 'Jeg liker svært godt å gå i selskaper hvor jeg møter mange mennesker'.» Begge metodene har funnet de samme fem store toppene i det personlighetsmessige terrenget – «The Big Five» (John, 1990; McCrae & Costa, 1997).

Femfaktormodellen og NEO PI-R

Femfaktormodellen (FFM) for personlighetstrekk postulerer at hver av de fem store personlighetsdimensjonen, kalt domener, består av seks underdimensjoner, kalt fasetter. Det blir dermed 30 små og fem store faktorer i terrenget av personlighetsegenskaper eller personlighetstrekk.

FFM har fått et internasjonalt standardinstrument for måling av de fem store og de 30 små faktorene, nemlig NEO PI-R (Revised NEO Personality Inventory) (Costa & McCrae, 1992). Dette instrumentet består av 240 utsagn, som alle beskriver en atferdstilbøyelighet, som personen skal uttrykke sin grad av enighet med på en femdelt skala fra 0 til 4, alt etter hvor godt personen synes utsagnet passer på ham/henne selv. Det er åtte utsagn for hver fasett som summeres til en råskåre for fasetten. Svarene på alle de seks fasettene som inngår i hvert domene, dvs. 48 svar på like mange utsagn, summeres til råskåren for hvert domene. Råskårene overføres til T-skårer for alle dimensjonene ved hjelp av normtabeller. T-skårene er tilnærmet normalfordelte med middelvei 50

og SD 10. Til slutt tegnes en profil over alle de 30 fasettene, og de fem domenene eller storfaktorene, hvor hver dimensjon betegnes med kun ett navn, dvs. navnet på et personlighetstrekk.

McCrae og Costa (1992) har benyttet personers enighet i den profilen som tegnes etter deres svar på NEO PI-R som en indikasjon på instrumentets begrepsvaliditet. Så vidt vi vet har kun én undersøkelse studert personers grad av enighet med en vilkårlig FFM-profil, altså en profil som ikke er deres egen, men hvor likheten med deres reelle profil varierer tilfeldig fra dimensjon til dimensjon (Andersen & Nordvik, 2002). Resultatene viste en klar tendens til at respondentene sa seg mer enige jo mer lik den tegnede skåre var deres reelle, dvs. skårene de fikk når de etterpå besvarte testen. Dette tyder på at den såkalte Barnum-effekten, dvs. tendensen til å godta ethvert utsagn som gjeldende for oss selv, i liten grad gjelder for personlighetsbegrepene i FFM. Når folk finner at horoskop i ukebladene stemmer, er det trolig fordi utsagnene er så generelle at de stemmer for alle.

Skåreprofilen fra NEO PI-R inviterer til å studere samsvar mellom de to variantene av selvrappport, nemlig den hvor personer vurderer seg selv direkte på personlighetsbeskrivende begreper, og den hvor personer vurderes seg selv indirekte, dvs. i forhold til utsagn om tilbøyeligheter, slik det benyttes i testleddene i NEO PI-R. Det var en slik sammenligning Andersen og Nordvik (2002) gjorde.

Vår hensikt var den samme, men vi benyttet en fremgangsmåte som utelukker Barnum-effekten. Personene fikk ingen ferdigtegnet profil; de ble bedt om selv å tegne sin egen profil på profilarket for NEO PI-R, slik de trodde den ville bli. Denne ble så sammenholdt med profilen basert på testresultatene.

Metode

Deltakerne i undersøkelsen var 40 studenter som fulgte kurset i personlighetspsykologi ved Psykologisk institutt, NTNU, våren 2003; 17 var cand.polit.-studenter og 23 var cand.psychol.-studenter. Det var 29 kvinner og 11 menn i utvalget. Alder varierte fra 23 til 39 år med middelverdi 26 år og SD 4.7 år.

Fremgangsmåte

Undersøkelsen ble gjort i tilknytning til den obligatoriske øvelsen med den norske oversettelsen av NEO PI-R (Martinsen, Nordvik, & Østbø, 2003, 2005) på personlighetskurset. Før studentene besvarte testen, fikk de utlevert profilarket og ble bedt om å tegne sin egen profil i form av T-skårer

på arket. Denne oppgaven krever at man kjenner begrepene i FFM og vet hva T-skårer er. Det var sikret ved at studentene på forhånd hadde gjennomgått statistikk og testmetode. De hadde umiddelbart før øvelsen deltatt i forelesninger om trekkteori og FFM, og hadde lest Costa og McCraes (1992) definisjoner og beskrivelse av alle dimensjonene i FFM i manualen for NEO PI-R.

Etter at de hadde tegnet sin profil, besvarte de testen og tegnet inn resultatet på det samme profilarket som de hadde forhåndstegnet sin profil på. Det ble presisert og kontrollert at endring av forhåndsprofilen når testresultatet av testen forelå, var strengt forbudt.

Resultater

Tabell 1 viser middelerverdier og *SD* for estimerte (forhåndstegnede) og oppnådde testskårer. Samsvar mellom estimert og oppnådd skåre er beregnet på to måter: For det første hvor godt gruppens gjennomsnittlige estimerte skåre samsvarer med gjennomsnittsskåren fra testen. For det andre hvor godt den enkelte persons estimerte skåre samsvarer med den oppnådde, uttrykt i korrelasjonen mellom estimert og oppnådd testskåre i alle variablene som blir målt med NEO PI-R.

Tabell 1. Middelerverdi (M) og standardavvik (SD) for testskårer og estimerte skårer samt differansen mellom middelerverdiene, differansens effektstørrelse (d) og korrelasjonen mellom testskårer og estimerte skårer.

Fasetter og domener	Testskåre		Estimert skåre		Differanse	d	Korrelasjon
N1: Angst	50.7	12.2	46.6	10.8	4.1	.4	.70 **
N2: Fiendtlighet	47.8	9.4	43.1	11.0	4.7	.5	.47 **
N3: Depresjon	49.6	9.5	46.9	11.3	2.7	.3	.66 **
N4: Selvbevissthet	52.8	9.6	52.8	10.1	0.0	.0	.44 **
N5: Impulsivitet	48.5	11.9	43.6	10.9	4.9	.4	.45 **
N6: Sårbarhet	49.8	11.1	50.0	11.5	0.2	.0	.67 **
N: Nevrotisme	50.2	10.0	48.8	9.5	1.4	.1	.76 **
E1: Varme	53.0	8.1	58.1	7.1	-5.1	-.7	.40 **

Fasetter og domener	Testskåre		Estimert skåre		Differanse	d	Korrelasjon
E2:							
Sosiabilitet	52.6	9.0	54.8	9.4	-2.2	-.2	.54 **
E3:							
Selvmarkering	50.8	11.2	52.8	9.4	-2.0	-.2	.73 **
E4: Aktivitet	49.5	10.5	53.0	10.8	-3.5	-.3	.64 **
E5:							
Spenningsøkning	51.3	8.4	50.3	9.8	1.0	.1	.54 **
E6: Positive følelser	50.3	9.5	55.8	8.6	-5.5	-.6	.58 **
E:							
Ekstroversjon	51.2	9.0	54.3	7.9	-3.1	-.4	.65 **
O1: Fantasi	54.0	9.6	56.1	10.9	-2.1	-.2	.73 **
O2: Estetikk	55.9	7.9	56.5	9.5	-0.6	-.1	.46 **
O3: Følelser	56.9	9.0	60.7	7.8	-3.8	-.5	.48 **
O4:							
Handlinger	54.4	9.8	55.2	7.8	-0.8	-.1	.49 **
O5: Ideer	55.0	10.2	57.6	9.4	-2.6	-.3	.37 *
O6: Verdier	55.7	8.0	57.6	9.1	-1.9	-.2	.02
O: Åpenhet	57.6	8.1	58.1	6.6	-0.5	-.1	.37 *
A1: Tillit	53.1	12.6	55.3	10.3	-2.2	-.2	.63 **
A2:							
Rettferdighet	51.1	12.5	53.2	9.8	-2.1	-.2	.36 *
A3:							
Altruisme	53.8	8.6	54.5	8.2	-0.7	-.1	-.15
A4:							
Føyerlighet	51.3	9.7	49.6	8.0	1.7	.2	.34 *
A5:							
Beskjedenhet	48.8	7.1	50.6	9.0	-1.8	-.2	.31
A6:							
Følsomhet	54.1	8.3	57.2	7.6	-3.1	-.4	.29
A:							
Medmenneskelighet	52.3	9.2	54.8	7.7	-2.5	-.3	.37*
C1:							
Kompetanse	52.2	9.8	56.2	7.3	-4.0	-.5	.41 **
C2: Orden	48.9	11.0	50.6	9.5	-1.7	-.2	.68 **
C3:							
Plikttroskap	50.8	8.2	57.1	8.6	-6.3	-.8	.71 **

Fasetter og domener	Testskåre	Estimert skåre	Differanse	<i>d</i>	Korrelasjon		
C4: Prestasjonsstreben	54.7	9.6	58.7	9.6	-7.0	-.7	.57 **
C5: Selvdisiplin	49.3	11.1	53.9	9.4	-4.6	-.5	.63 **
C6: Betenksomhet	52.6	7.6	57.8	7.5	-5.2	-.7	.38 *
C: Planmessighet	51.2	10.0	56.0	6.5	-4.8	-.6	.63 **

Det er på domenet Åpenhet med tilhørende fasetter at begge sett av skårer er høyest og spredningen minst. Høy Åpenhet var det som sterkest karakteriserer studentgruppen sammenlignet med normeringsutvalget. Det samme, men i mindre grad, gjelder for Medmenneskelighet.

Ser vi på retningen av forskjellene i middelværdier, er det 26 som er negative. Dette innebærer at den estimerte skåre var høyere enn testskåren, mens ni er positive, dvs. at testskåren var høyere enn den estimerte. Sju av de ni finner vi i Nevrotisme, hvor så vel totalskåren som alle fasettene har en positiv forskjell. Det er altså en tendens til at studentene underestimerte sine skårer på nevroisismefasettene i motsetning til de øvrige dimensjonene, hvor studentene ganske konsekvent overestimerte sine skårer.

Cohens *d* er standardmål for størrelsen av forskjell mellom middelværdier, hvor differansen måles med gruppenes felles («pooled») *SD* som enhet (1992). En *d* lik .2, .5 og .8 regnes som henholdsvis liten, middels og stor effektstørrelse (Cohen, 1992). Det er på domenet Planmessighet studentene mest overvurderer sin skåre, og fasetten O4: Plikttroskap er den eneste hvor effektstørrelsen er stor ut fra Cohens kriterium. De fleste forskjellene mellom den virkelige og den estimerte skåren er ikke store. Grovt sett, på gruppenivå, ga studentene derfor et godt estimat av sin testskåre.

Alle korrelasjonene mellom estimert og oppnådd skåre var positive, unntatt en. Det var kun på fire av fasettene at korrelasjonen mellom estimert og oppnådd skåre ikke var statistisk signifikant ($p < .05$). Disse fasettene og korrelasjonene var O6: Verdier (.02), A3: Altruisme (-.15), A5: Beskjedenhet (.31), og A6: Følsomhet (.29). På individnivå var det altså en klar tendens til at studentene estimerte sin skåre langt bedre enn det tilfeldige. Den høyeste korrelasjonen (.76) var på Nevrotisme. Selv om dette domenet var det som studentene svakt, men konsekvent underestimerte sin skåre på, var det her at de mest konsekvent korrekt estimerte sin relative posisjon,

sammenlignet med andre. Den eneste negative korrelasjon gjaldt for fasetten Altruisme under domenet Medmenneskelighet. Det betyr at det var på denne fasetten at studentene mest feilestimerte sin skåre.

Tabell 2 viser korrelasjonen mellom den reelle skåren på de fem domenene og den absolutte differansen mellom estimert verdi og testskåren. En positiv korrelasjon betyr at jo høyere testskåren var desto større var differansen mellom estimert og oppnådd skåre, uansett retningen på differansen. Differanseskårer har lav reliabilitet; jo høyere korrelasjonen mellom to variabler er desto mindre reliabel blir differanseskårene. At alle korrelasjonene i Tabell 2 er lave er derfor i betydelig grad en nødvendig konsekvens av de høye korrelasjonene i Tabell 1. Spesielt gjelder dette for korrelasjonene med differanseskåren i Nevrotisisme i øverste rad av tabellen pga. den høye korrelasjonen mellom estimert verdi og testskåre i dette domenet (Tabell 1). Men det faktum at alle de fem korrelasjonene for skåren i Nevrotisisme er positive, mens 18 av de øvrige 20 er negative, er ikke statistisk tilfeldig. Det viser en tendens til at høy Nevrotisisme var svakt forbundet med feil i estimering for alle domenene, mens høy skåre på de øvrige domenene var svakt forbundet med god estimering.

Tabell 2. Korrelasjon mellom testskåre og absolutt differanse mellom testskåre og estimert verdi.

Absolutt differanse	Testskåre				
N: Nevrotisisme	.11	-.13	.02	-.01	.06
E: Ekstroversjon	.03	-.38*	-.16	-.40*	-.25
O: Åpenhet	.08	-.09	-.11	-.22	-.17
A:					
Medmenneskelighet	.18	-.13	-.34*	-.29	-.31
C: Planmessighet	.21	-.29	-.14	-.08	-.37*

Diskusjon

Denne undersøkelsen hadde først og fremst en pedagogisk hensikt; vi ønsket å stimulere til forståelse og diskusjon av FFMs begreper og metoder for å måle dem. Vitenskapelig sett hadde vi ingen hypoteser, men det var noen spørsmål undersøkelsen kunne belyse.

I normeringsutvalget for NEO PI-R, som består av ca. 3500 personer, er middelveiden 50 og SD 10 for alle fasettene og domenene (Martinsen et al., 2003). Det mest konsekvente avvik

fra disse verdiene i studentgruppens skårer var på Åpenhet og fasettene under dette domenet, hvor middelverdiene var høye og standardavvikene små. Tilsvarende gjelder, men i mindre grad, for Medmenneskelighet. Skal vi snakke om en psykologprofil, må det bli at den består av høy Åpenhet og Medmenneskelighet.

Det domenet med tilhørende fasetter hvor studentgruppen mest konsekvent overvurderte sin skåre, var Planmessighet. M og SD i testskårene er 51 og 10, altså svært nær populasjonsverdiene, men de estimerte verdiene var 56 og 6.5. Studentene tror de er mer planmessige enn folk flest; det er de enige om, men de tar feil om vi skal tro på testresultatet hvor deres svar sammenlignes med «folk flest».

Hovedspørsmålet

Det overordnede spørsmålet var om studentene kunne bestemme sin testskåre på forhånd, ut fra kjennskapet til seg selv og til trekkbegrepene i FFM. Svaret er at studentene gjorde dette mye bedre enn en tilfeldig gjetning. Det var bare fire av korrelasjonene mellom estimert og oppnådd skåre som ikke var statistisk signifikante. Det gjaldt for fasettene O6: Verdier, A3: Altruisme, A5: Beskjedenhet og A6: Følsomhet. Det kan virke som disse fasettene har noe felles. De tre A-fasettene peker alle mot et dypt og stille engasjement for medmennesker. O6: Verdier står for toleranse for sosiale, kulturelle og religiøse verdier, som peker mot en «intellektuell» medmenneskelighet, som lar seg forene med de nevnte A-fasettene. At disse fasettene skulle være de som studentene dårligst estimerte sine skårer på, hadde vi ingen forhåndshypotese om. Etterpå kan vi kanskje konstatere at de som er besjelet med beskjeden uselvishet og omtanke for andre, ikke ser det selv; de merker i hvert fall ikke av dette når de tegner sin profil. Er det slik at snille, uselviske og tolerante mennesker ikke er seg bevisst at de er slik, eller er de så beskjedne at de ikke vil vedstå seg det?

Reliabilitet

Korrelasjoner mellom målte variabler begrenses av reliabiliteten i målingene. Reliabiliteten i skårene på NEO PI-R, uttrykt i alfakoeffisienter, varierer for fasettene fra .54 for C1: Kompetanse, til .84 for N3: Depresjon. For domenene varierer koeffisientene fra .84 for Medmenneskelighet, til .92 for Nevrotisisme (Martinsen et al., 2003). Korrelasjoner mellom to skalaer som har samme sanne skårer og samme reliabilitet, vil være lik med skalaenes reliabilitet. To variabler kan ikke korrelere høyt med mindre de har høy reliabilitet (Nordvik & Ulleberg, 2000). Vi kjenner ikke reliabiliteten i de estimerte verdiene, men siden de ikke kan forventes å korrelere med andre variabler med mindre de har en relativt høy reliabilitet, blir de høye korrelasjonene i Tabell 1 evidens for at de estimerte verdiene

gjennomgående måtte være reliable. Spørsmålet er om de få lave koeffisientene kan komme av at de aktuelle estimerte verdiene hadde lav reliabilitet. Det vet vi ikke noe om, men det virker usannsynlig at reliabiliteten skulle variere så mye og så systematisk treffe variabler som hadde et felles innhold som antydnet ovenfor.

Hva om vi hadde funnet en perfekt korrelasjon mellom estimert og oppnådd skåre på NEO PI-R? Det måtte bety at testen ble overflødig, i hvert fall for psykologistudenter; det ville være nok å la studentene tegne sin egen profil uten den tidkrevende besvarelsen av alle leddene i testen. Men slik er det ikke. Om det er studentenes tegninger av sin personlighetsprofil eller den profilen som resultatene fra NEO PI-R tegner, som er mest korrekt, kan vi ikke avgjøre, men faktum er at de ikke er identiske, selv om de er korrelerte.

Validitet

Et spørsmål som ofte stilles i forbindelse med personlighetstester av selvrapporttypen, er om det ikke er lett å «gjennomskue» og svare slik at resultatene blir slik man ønsker, eller slik man antar er ønskelig. McCrae og Costa (1997) har argumentert både mot at dette er mulig i særlig grad, og at personer prøver å gjøre det. Våre resultater støtter dette synet. I undersøkelsen ble det gjort eksplisitt at det dreide seg om to metoder for å måle det samme: først en direkte selvvurdering på dimensjonene i FFM og dernest selvvurdering på de 240 leddene i NEO PI-R. Studentene hadde sin tegnede profil tilgjengelig mens de besvarte testen; de ble verken frarådet eller tilrådet å konferere med profilen mens de besvarte testen. Det kunne tenkes at denne situasjonen inviterte til å svare slik at resultatet ville samsvare maksimalt med den selvbeskrivelse som allerede var gitt. Det skjedde ikke; alle rapporterte etterpå at dette var de overhodet ikke opptatt av. De formidlet at de ble helt konsentrerte om det enkelte testledd og å vurdere seg selv mest mulig korrekt i forhold til det. Dette er et forsvar for den umiddelbare validiteten i selvrapport av den typen som benyttes i NEO PI-R. Hvor generelt det er at personer blir konsentrert om testleddene og «glemmer» andre hensyn, vet vi ikke, men det virker ikke sannsynlig at psykologistudenter er spesielle i denne henseende.

At det er et moderat samsvar mellom resultatene fra de to metodene, indikerer at de har en betydelig felles begrepsvaliditet. At samsvaret ikke er perfekt, indikerer at å vurdere seg selv på en trekkbetegnelse, ut fra den forståelse man har for betegnelsen og de assosiasjoner den gir, ikke er helt det samme som å vurdere seg selv på de beskrivelser av atferdstilbøyeligheter som NEO PI-R inneholder. Dette kan utgjøre et forsvar for validiteten i NEO PI-R; ingen, ikke en gang psykologistudenter, kan bestemme hvor ekstroverte de er; de må bekrefte en rekke utsagn om sine

tilbøyeligheter for at dette kan bestemmes. Gyldigheten av denne påstanden er imidlertid avhengig av hvem som vet best om ekstroversjon og andre trekk, de som har laget utsagnene i NEO PI-R eller «folk», inkludert psykologistudenter.

Avslutning

Vi har i denne artikkelen holdt oss strengt til det deskriptive, hva data viser, og at dataene kommer fra psykologistudenter. Generaliserbarheten er begrenset, om ikke til psykologistudenter, så til personer som kjenner personlighetspsykologiske begreper i sin alminnelighet og begrepene i FFM i særdeleshet. Å bruke testledd, som i NEO P-R, er i prinsippet en rasjonell erstatning for atferdsobservasjon; i stedet for å observere hvordan folk ter seg, blir de spurt om det. Når svar på bestemte kategorier av testledd summeres, får man en faktorskåre. Den kan tolkes rent deskriptivt, som en oppsummering av selvrapportert atferd, men ofte tolkes den som et mål på personens plass på en latent dimensjon, et trekk, som manifesteres i atferd. Det er når dette trekket gis et navn, at muligheten oppstår for å la personer vurdere seg direkte på det navnet betegner, slik de oppfatter det. Det krever åpenbart mindre psykologisk begrepskunnskap å besvare testledd enn å vurdere seg selv på ett trekknavn. Det er derfor personlighetstester som NEO PI-R kan brukes på personer som ikke er skolerte i personlighetspsykologi. At personer som er skolerte, som psykologistudenter, i stor grad, men på ingen måte helt konsekvent, gir selvvurderinger på trekkbegrepene som samsvarer med testskårene, er vel omtrent som forventet.

Hilmar Nordvik

Psykologisk institutt, NTNU

7491 Trondheim

Tlf 73 59 19 77

E-post hilmar.nordvik@svt.ntnu.no

Referanser

Andersen, P., & Nordvik, H. (2002). Possible Barnum effect in the five factor model: Do respondents accept random NEO Personality Inventory-Revised scores as their actual trait profile? *Psychological Reports, 90*, 539–545.

- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Costa, Jr., P. T. Jr., & McCrae, R. R. (1992). NEO PI-R. *Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, Jr., P. T. Jr., & McCrae, R. R. (1997). Stability and change in personality assessment: The revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment*, 68, 86–94.
- John, O. (1990). The «Big Five» factor taxonomy: Dimensions of personality in the natural language and questionnaires. I L. A. Pervin (Ed.), *Handbook of personality. Theory and research* (ss. 66–100). New York: Guilford Press.
- Martinsen, Ø., Nordvik, H., & Østbø, L. E. (2003). *Norsk utgave av Revised NEO Personality Inventory (NEO PI-R)*. Oslo: Gyldendal Akademisk.
- Martinsen, Ø. L., Nordvik, H., & Østbø, L. (2005). Norske versjoner av NEO PI-R og NEO FFI. *Tidsskrift for Norsk Psykologforening*, 42, 421–423.
- McCrae, R. R., & Costa, P. T. Jr. (1995). Trait explanations in personality psychology. *European Journal of Personality*, 9, 231–252.
- McCrae, R. R., & Costa, P. T. Jr. (1997). Personality trait structure as a human universal. *Psychological Bulletin*, 52, 509–516.
- Nordvik, H., & Ulleberg, P. (2000). *Teststatistikk*. Trondheim: Tapir Akademisk.