

## Methodologie TOP-test bekeken

Bertus Jeronimus

### Introductie

De introductie van de handleiding is informatief en rijk in voorbeelden maar mist soms ook nuance en lijkt met vlagen uit de AI getrokken. De kernargumentatie is dat persoonlijkheidsonderzoek in de context van werk vaak betrekking heeft op de 'lichte kant' oftewel positieve eigenschappen van persoonlijkheid en niet naar de 'donkere kant' oftewel ongewenst gedrag op werk.

De meeste variantie in de *Dark Triad* is al gevangen door altruïsme maar daar wil men niet aan in de handleiding. Op zijn best zal de *Dark Triad*-vragenlijst meer details of een meer ernstige verschijningsvormen meten van dezelfde dimensionale onderliggende duistere eigenschappen. De claim dat de TOP vereist is voor het in kaart brengen van donkere persoonlijkheidseigenschappen bij personeelsselectie lijkt me dus onjuist, dat kan al met uitgebreide Big Five-schalen. Er is geen discussie van de verschillen of bewijs voor incrementele voorspellende capaciteiten.

Het verkooppunt van de TOP is het gebruik van een vergevingsgezind of meer acceptabel vocabulaire om makkelijker het gesprek met werknemers aan te kunnen gaan en de directe (nauwere) toepassing op de werkcontext. Met de *Dark Triad of Personality at Work* (TOP) claimt men dat er nu een wetenschappelijk onderbouwd instrument is dat specifiek gericht is op het in kaart brengen van narcisme, machiavellisme en psychopathie in een werkgerelateerde context. Die waren er al.

Er wordt gesteld dat er geen geschikte Nederlandse alternatieven zijn, maar dat is gewoon niet waar, er zijn meerdere instrumenten die goed zijn onderzocht en gevalideerd en in combinatie met de Big Five effectief zijn in het meten van de *Dark Triad* en gerelateerd werkgedrag. De Dark-Triad DirtyDozen (DTDD) is in het Nederlands goed onderzocht (e.g., Klimstra et al., 2014) en er zijn allerlei versie van de Narcissistic Personality Inventory (NPI), de Machiavellianismtest (MACH-IV) en Levensons psychopathieschalen (LSRP) zijn beschikbaar die veel worden gebruikt in organisaties 'voor de donkere kant van persoonlijkheid', net als de Hogan Development Survey (HDS). Daarnaast kan je in Nederland uitstekend begripsvaliditeit doen met bijvoorbeeld Big Five vragenlijsten.

### Beschrijving en rationale van de TOP schalen

De schalen hebben verschillende aantallen items (e.g., zelfverzekerde werkhouding, van 3 tot 9 items per subschaal). De items van de Nederlandse vertaling van het van oorsprong Duitse instrument zijn vertaald uit de Engelse versie, zonder argumentatie (waarom niet de Duitse?).

Bij de ontwikkeling van het instrument zijn de schalen binnen ieder van de drie trekken uitgewerkt, zonder na te denken over hun potentiële overlap (die is artificieel 'opgelost' met het verwijderen van items met hoge ladingen op meer dan één van de factoren – terwijl deze overlap ingebakken zit in hun ontwerp). Drie van de elf zo ontwikkelde schalen bleken op een ander bovenliggende factor te laden dan dat ze ontworpen waren (e.g. sociale invloed, risicobereidheid en ongevoeligheid waren ontworpen voor psychopathie, maar de eerste twee bleken op narcisme te laden). De daaropvolgende verklaringen voor deze verschillen is post-hoc klets. 'Al met al moet gesteld worden dat de TOP aansluit bij de klassieke

benadering van de *dark triad*, waarbij wordt uitgegaan van aparte constructen die voor een bepaald deel met elkaar overlappen.' Heel vreemd om dan eerst items met overlap te verwijderen, als daarna het laden van schalen op andere factoren als onderbouwing voor deze conclusie moet doorgaan.

De auteurs maakten dertien nieuwe items die niet in het oorspronkelijke instrument zaten, en hebben zes items met matige ladingen vervangen door zes uit deze lijst van dertien. Het blijft onduidelijk waarom, en het ontbreekt ook aan argumentatie voor de verschillen tussen deze items en de beter functionerende items (Tabel 5).

### **Normgroepen**

De normgroepen zijn niet representatief. Voor de scheve man-vrouwverdeling van leidinggevend en wordt verwezen naar internationale cijfers in plaats van een vergelijking met Nederlandse gegevens (midden- en hoger management ~30-33% vrouw – in lijn met de verdeling van Panelinzicht). Suggereert ook AI. Er zitten in de normgroepen mensen die minimaal 2 dagen per week werken, wat een nogal lage minimumgrens is, en niet onderbouwd, en meer pragmatisch ingestoken lijkt (zo werkt Panel-inzicht waar ze gebruik van maken). Bijna een kwart van de leidinggevend is zelfstandig ondernemer, en de meerderheid van de leidinggevend had <10 werknemers in hun team, dus dat kan ook 2 zijn. En bijna een vijfde van de niet-leidinggevend was een zelfstandige zonder personeel, dus ook eigen baas. Niet alleen betreft dit dus steeds een hele specifieke groep mensen (dus waarschijnlijk niet normatief voor *Dark Triad*), ze hebben ook niet of anders te maken met leidinggevend dan de modale werknemer.

De handleiding beschrijft de interne consistentie uitgebreid, maar de auteurs bedoelen hier eigenlijk de betrouwbaarheid (Cronbachs alfa is geen maat voor de interne consistentie (homogeniteit van vragen) maar meet hoe sterk items in de schaal met elkaar correleren, een ondergrens van de betrouwbaarheid). Het lijkt erop dat dit instrument ook wordt aanbevolen voor selectie. Daarvoor gelden hele strenge regels als het gaat om de betrouwbaarheid. Een waarde van .80 of hoger is voldoende, een waarde van .90 of hoger goed. De gemiddelde betrouwbaarheidscoefficiënt van de schalen is .82 bij de leidinggevend en .80 bij de niet-leidinggevend. De stabiliteit van de TOP is het over het algemeen matig te noemen met een gemiddelde intraklassecorrelatiecoëfficiënt van .63 voor de factoren en .59 voor de schalen. Dit betekent dat er relatief veel variatie zit tussen de antwoorden op de twee meetmomenten, waardoor de scores slechts matig overeenstemmen.

Alle schalen werden ingezet en niet alleen de drie hoofdschalen (overigens is mij nog steeds niet duidelijk waar al die subschalen vandaag komen). Als ik naar de betrouwbaarheidsschattingen kijk, zie ik een waarde van .44 en .61 voor Flexibiliteit. Ook de overkoepelende schaal waar dit in valt heeft een onvoldoende betrouwbaarheid bij niet-leidinggevend. Dus diverse subschalen zijn onvoldoende als dit ook voor selectie wordt ingezet, of zijn sowieso onvoldoende. Ze geven gelukkig zelf al aan dat de stabiliteit matig is. Qua begripsvaliditeit is er een CFA gedaan op de gecombineerde normgroepen, die slechte uitkomsten geeft. Het blijft onduidelijk waarom het model met drie factoren superieur zou zijn.

Er is geen onderzoek gedaan naar de verbanden tussen de TOP en andere instrumenten in Nederland (convergente validiteit). Onderzoek uit andere landen wordt genoemd, maar is niet relevant. De studie met de NEO-PI-3 is misleidend omdat niet het hele instrument en spectrum is afgenomen, om de incrementele capaciteiten van de TOP ten opzichte van de

zes facetten in de Altruïsmefactor te laten zien, bijvoorbeeld. En zeker ook als Consciëntieusheid en Neuroticisme mee worden genomen. Een gemiste kans. Ook zijn de 108 leidinggevenden en 93 werknemers die de NEO hebben ingevuld onvoldoende voor betrouwbare correlaties waar het minimumaantal van 250 doorgaans als ondergrens wordt gezien. Daar voldoet de hele steekproef van 117 mannen en 84 vrouwen niet eens aan. Dit is dus hetzelfde voor de vergelijkingen met andere *Dark Triad*-lijsten en de intelligentietestscores.

### **Conclusie**

Na het doorlezen zou ik de huidige uitgangspunten, normen, betrouwbaarheid (van enkele subschalen), begripsvaliditeit en criteriumvaliditeit als matig tot zwak beschrijven. De kwaliteit van de handleiding en het testmateriaal is al een stuk beter.

### **Wat is er goed?**

Dat er wel zorg is voor de handleiding en zo, met mooie introductie en beschrijving van de literatuur, meer dan voor de feitelijke validatie eigenlijk.

Ze doen ook geen test-theoretische meet-invariantietoetsen, zoals dat de items hetzelfde functioneren voor mannen en vrouwen, of voor jongeren of ouderen - zoals wij eens deden voor een Dark Triad-vragenlijst in Klimstra et al. 2020. Dat is wel het minimum eigenlijk, zeker als de lijst nieuw is.

Klimstra, T. A., Jeronimus, B. F., Sijtsma, J. J., & Denissen, J. J. (2020). The unfolding dark side: Age trends in dark personality features. *Journal of Research in Personality*, 85, 103915.