

*Psychodiagnostiek behoort tot de kroonjuwelen van ons vak, schrijven Harald Merckelbach en Brechje Dandachi-FitzGerald.*

*Des te meer reden om eens stil te staan bij de valkuilen waarin de psychodiagnosticus kan tuimelen. Zoals de neiging om opdrachtgevers of verwijzers ter wille te zijn. Over deze en andere valkuilen gaat dit artikel. ‘Goede diagnostiek vereist tenminste dat een psycholoog het met zichzelf en met competente vakgenoten eens is.’*

## OBSTAKELS VOOR DE DIAGNOSTICUS

# PSYCHO- DIAGNOSTIEK ANNO 2020

Is kandidaat X geschikt voor deze baan? Heeft leerling Y faalangst? En lijdt patiënt Z aan een depressie? Met psychodiagnostiek laten zich zulke vragen beantwoorden. Psychodiagnostiek wordt hier in ruime zin begrepen: alles wat psychologen uit hun gereedschapskist halen om tot gefundeerde uitspraken over personen te komen. Zo bezien behoort psychodiagnostiek tot de kern van de psychologie. Ook al omdat men in het maatschappelijk verkeer voortdurend van psychologen verlangt dat zij bruikbare en gezaghebbende oordelen vellen: over sollicitanten, leerlingen, patiënten, justitiabelen en nog zo wat groepen. Psychologen lijken vaak goed aan die wens te kunnen voldoen.

Het zou ook eenvoudig zijn om hier nu een catalogus van psychodiagnostische triomfen te ontvouwen. Daarin past bijvoorbeeld het relaas van de jongen die in een instelling voor kinderen met een verstandelijke beperking werd geplaatst, maar vervolgens op een IQ-test 108 scoorde. Waarna hij de instelling kon verlaten en in de gelegenheid werd gesteld om een middelbare beroepsopleiding te volgen.<sup>1</sup> Maar ook het onderzoek dat liet zien dat je met pre-chirurgische psychodiagnostiek in 80% van de gevallen kunt

voorspellen of een patiënt baat gaat hebben bij een hernia-operatie of niet (Block et al., 2001). En voorts het inzicht dat behandelingen effectiever worden als psychotherapeuten hun patiënten regelmatig aan diagnostiek (laten) onderwerpen en zich van de uitkomsten daarvan op de hoogte stellen (Scott & Lewis, 2015).

In dit artikel willen we niet zozeer stilstaan bij psychodiagnostische succesnummers, maar bij obstakels die eraan in de weg staan. Daarbij valt bijvoorbeeld te denken aan de neiging van de diagnosticus om de opdrachtgever of verwijzer ter wille te zijn. Dat is een valkuil en de diagnosticus die erin tuimelt, zal tekort schieten. Hieronder zetten we meer van zulke valkuilen op een rijtje. We leggen ook uit hoe ze te ontlopen zijn; bij wijze van voorschot attenderen we alvast op tabel 1. Voordat we daaraan toekomen eerst deze vraag: hoe weten we wanneer psychodiagnostiek tekort schiet? Meestal ontbreekt een gouden standaard – er zijn geen onafhankelijke criteria voorhanden om vast te stellen dat de sollicitant geschikt is, de patiënt depressief en zo meer.<sup>2</sup> Laat staan dat we eenduidig zouden kunnen bepalen

1 Zie: Merel van Leeuwen. Gevangen tussen zwakbegaafden. *De Limburger* 18 februari 2014.

2 Zouden zulke gouden standaards er wel zijn, was psychodiagnostiek overbodig. Het is precies de afwezigheid van obligate standaards die de validiteit van psychodiagnostiek tot zo'n lastig thema maakt.

dat de psycholoog er met zijn diagnostische uitspraken faliekant naast zat.

### DIAGNOSTISCHE LABILITEIT

Toch valt er meer over te zeggen. Goede diagnostiek vereist tenminste dat een psycholoog het met zichzelf en met competente vakgenoten eens is. Dus diagnostiek faalt als een psycholoog bij mensen met dezelfde kenmerken de ene keer, pakweg, een depressie en de volgende keer een angststoornis vaststelt. De consistentie – uitgedrukt in de correlatiecoëfficiënt ( $r$ ) – laat dan te wensen over. Diagnostiek faalt evenzeer wanneer – om bij het voorbeeld te blijven – eenzelfde persoon door de ene vakgenoot met een depressie en door de andere (even kundige) vakgenoot met een angststoornis wordt gediagnosticeerd. De consensus – uitgedrukt in kappa ( $k$ ) – is dan bescheiden.<sup>3</sup> Over de volle breedte van hun vak beschouwd – dus ruimer dan psychodiagnostiek – zijn consistentie en consensus van psychologen niet direct om over te jubelen. Met hun geschatte consistentie en consensus van respectievelijk 0.44 en 0.40 doen ze het weliswaar beter dan effectenhandelaren (wier uitspraken over verwachte koersstijgingen niet te onderscheiden zijn van koffiedik kijken), maar wezenlijk slechter dan bijvoorbeeld meteorologen (wier voorspellingen over het weer van morgen tamelijk nauwkeurig blijken; Thomas & Lawrence, 2018<sup>4</sup>).

Gebrekkige consistentie en consensus zijn aan de orde wanneer patiënten met psychische problemen wisselende diagnoses te horen krijgen als ze herhaaldelijk worden onderzocht. Australische onderzoekers volgden bijvoorbeeld 700 forensische patiënten en vonden dat van de patiënten die aanvankelijk met een angst- of stemmingsstoornis werden gediagnosticeerd minder dan de helft bij heronderzoek dezelfde diagnose kreeg (Schilders & Ogloff, 2017). Dat type diagnostische labiliteit werd ook beschreven in een grootschalig onderzoek onder Spaanse patiënten (Baca-Garcia et al., 2007; zie voor andere voorbeelden: Mokros, Habermeyer & Küchenhoff, 2018).

In ons land was er het herdiagnose-project van de Gelderse zorgaanbieder GGNet onder 967 patiënten

(Timmermans, 2018). Er werd opnieuw gepraat met de patiënten en hun dossiers werden tegen het licht gehouden. Bij 257 patiënten (26%) achtten de experts de eerdere diagnose niet (langer) juist of volledig. Nog een opvallend cijfer: 165 mensen (17%) werden per direct uit de behandeling ontslagen, klaarblijkelijk omdat ze niet (langer) in aanmerking kwamen voor een diagnose. Het zijn cijfers die vragen oproepen over hoe psychologen te werk gaan als zij diagnosticeren. Welke obstakels komen ze tegen en onderkennen ze die ten volle (zie ook tabel 1)?

### CONTEXT

Een eerste obstakel kan zich voordoen zodra de psycholoog kennis neemt van de opdracht. In zo'n opdracht ligt een vraag besloten en die staat niet op zichzelf. Ze komt van een opdrachtgever of verwijzer. Dat is een bedrijf op zoek naar een geschikte werknemer, een huisarts die wil weten of een patiënt ADHD heeft, een rechter-commissaris die geïnteresseerd is in de persoon van de verdachte en zo zijn er legio andere voorbeelden van opdrachtgevers of verwijzers die allemaal zo hun verwachtingen hebben.

In het kielzog daarvan duiken onherroepelijk documenten en bijlagen op. Sommige daarvan zijn informeel. Neem *e-mails* van het type: 'Zelf hebben we twijfels over of deze kandidaat wel zelfstandig genoeg kan werken.' Of: 'De verdachte was eerder in behandeling vanwege een persoonlijkheidsstoornis.' Andere documenten zijn vakinhoudelijk van aard. Neem een verslag van een hersenscan met als conclusie: 'Patiënt heeft fronto-temporale beschadigingen.'

Op grond van de literatuur over cognitieve valkuilen (*biases*) weten we dat de context van opdrachtgevers en begeleidende documenten de diagnostische oordeelsvorming zwaar kan vertekenen (Wedding & Faust, 1989; Richards et al., 2015). Berucht is het fenomeen van diagnostisch momentum; de verwijzer meldt aan de psycholoog welke diagnose hij zelf in gedachte heeft en de psycholoog neemt dat als startpunt en gaat op zoek naar verdere onderbouwing. Dan dreigt het gevaar van tunnelvisie. Zo gaven de Nijmeegse onderzoekster Nanon Spanjaars en haar collega's (2015) ervaren en minder ervaren psychologen een dossier van een depressieve patiënt ter beoordeling. Of de psychologen maar een diagnose wilden stellen. Sommigen kregen van tevoren een verwijfsbrief te lezen van een huisarts over dezelfde patiënt. In deze verwijfsbrief werd ten onrechte gemeld dat de patiënt aan een angststoornis leed. Van de minder ervaren psychologen liet 55% zich meeslepen door de misleidende informatie, terwijl dit percentage bij de ervaren

3 Zowel consistentie ( $r$ ) als consensus ( $k$ ) variëren tussen de 0 (hoogst inconsistent en geen consensus) en de 1 (perfecte consistentie, volledige consensus). De opvattingen over afkappunten voor aanvaardbare consistentie en consensus lopen uiteen, maar vaak wordt de ondergrens gelegd bij  $r > 0.70$  en  $k > 0.40$ .

4 Consistentie en consensus van effectenhandelaren:  $< .40$  en  $0.32$ . Die van meteorologen:  $0.98$  en  $0.95$ . Overigens: de statistieken van Kraemer et al. (2012) over consistentie en consensus van medici en hun diagnoses zijn iets, maar niet heel veel beter dan die van psychologen.

TABEL 1. PSYCHODIAGNOSTIEK: PROBLEMEN EN WAT ERAAN TE DOEN VALT

INFORMATIEBRON	PROBLEMEN	OPLOSSING
Context	Diagnostisch momentum Affiliatie bias	Protocollen Alternatieve hypothesen Blinding voor dossierinformatie
Observatie	Overwaardering saillante details	Hypothese-toetsing
Interviews	Stereotypen Base rate neglect	Rijke prototypes Kennis van epidemiologie
Tests	Preoccupatie met uitbijters Slordige testattitude, veinzen of mooipraterij Gebrekkig zelfinzicht	ANDI Symptoomvaliditeitstests Informatie van derden

groep nog altijd 40% bedroeg. De studie van Spanjaars en collega's laat zien hoe terloopse informatie als anker kan gaan fungeren. Zoiets zal de consistentie en consensus niet ten goede komen.

Minstens zo belangrijk is wat in de literatuur wordt aangeduid als de *affiliatie-bias*. Die komt erop neer dat deskundigen hun opdrachtgever te vriend willen houden. In een fraai opgezet experiment benaderden de Amerikaanse rechtspsychologen Murrie en Boccaccini (2015) een groep collega's met het verzoek of ze voor 400 dollar wilden kijken naar de justitiële stukken van een viertal verdachten om daarna een oordeel uit te spreken over de mate waarin zij – de verdachten – psychopathische trekken vertoonden. De collega's die hier voor in waren, hadden een ontmoeting met de opdrachtgever. De ene keer deed deze zich voor als openbare aanklager; de andere keer als advocaat van de verdachten. In alle gevallen maakte de opdrachtgever toespelingen op toekomstige opdrachten. In alle gevallen ook bekeken de psychologen dezelfde vier dossiers. De psychologen die voor de openbare aanklager dachten te werken schatten de verdachten als aanmerkelijk gevaarlijker in dan de psychologen die voor de advocaat meenden te werken. De consensus was dus ver te zoeken.

Protocollair werken is de manier bij uitstek om diagnostisch momentum en *affiliatie-bias* tegen te gaan. In dit verband betekent het dat de psycholoog met opdrachtgever en onderzochte in gesprek gaat over de verwijsvraag en beziet of er alternatieve vragen onder de loep moeten worden genomen. Het betekent ook dat de psycholoog pas na

het psychodiagnostisch onderzoek kennis neemt van het volledige dossier van de onderzochte. En het betekent tenslotte dat de psycholoog zich bezint op de vraag of loyaliteit jegens de opdrachtgever een sturende invloed heeft (gehad) op de oordeelsvorming.

We stellen deze punten met nadruk aan de orde omdat bijvoorbeeld de geestelijke gezondheidszorg in ons land langs de lijnen van specifieke zorgprogramma's – aparte voorzieningen voor angststoornissen, ADHD, trauma - is georganiseerd. Die stoornisgerichte aanpak is een ideale voedingsbodem voor diagnostische verkokering (Eurelings & Van Os, 2018). Voorts: het advies om toch vooral vooraf kennis te nemen van het dossier behoort tot intellectuele inboedel van menig handboek.<sup>5</sup> Toch is er, zoals we hierboven hebben betoogd, alle reden om aan de wijsheid van dat advies te twijfelen.

### OBSERVATIE

Dus de psycholoog heeft kennis genomen van de verwijsvraag en met de onderzochte wordt een afspraak gearrangeerd, waarbij die om te beginnen nauwgezet wordt geobserveerd. Hier dient zich een volgende valkuil aan, ofschoon niet elke expert dat zo zal zien. Want sommige clinici

5 Een handboek over psychologische diagnostiek in de gezondheidszorg (Luteijn & Barelds, 2018, p. 25) zegt bijvoorbeeld: 'Bij de analyse van de aanmelding benut de diagnosticus dossiergegevens, zoals verslagen van eerder psychodiagnostisch of medisch onderzoek, en informatie uit bronnen zoals school, werk, gezin en instituten (zoals de rechtbank).'

menen dat je met enkel observeren heel ver kunt komen. Dat is tenminste de conclusie die zich opdringt uit onderzoek van de Duitse socioloog Von dem Knesebeck en collega's (2010). Ze lieten 128 artsen met een eigen praktijk (*Allgemeinmedizinern*) naar een korte video van een acteur kijken die geïnstrueerd was om over allerlei klachten te praten, voornamelijk vermoeidheid, gebrek aan energie, vage pijn, maar ook reumatische en cardiovasculaire symptomen. Een ruime meerderheid van de artsen (87%) vond de observatie voldoende om een depressie vast te kunnen stellen. Zo'n een op de vier artsen zou medicatie aan de acteur voorschrijven. Een minderheid – 20% – had behoefte aan nader testonderzoek. De observaties gaven, om kort te gaan, aanleiding tot overhaaste interpretaties.

Zijn psychologen wat dat betreft terughoudender? Ons is daarover geen geruststellend onderzoek bekend. Wel is er dat tweedelige BBC-programma uit 2008, getiteld *How mad are you?* (Proglora, 2009). Daarin mochten drie deskundigen, onder wie een hoogleraar klinische psychologie, tien kandidaten dagenlang observeren terwijl zij allerlei dingen deden, variërend van de *Wisconsin Card Sorting Test* tot het afvoeren van stalmest. Bij vijf van de tien was eerder vastgesteld dat ze kampten met ernstige psychische problemen zoals schizofrenie. Na dagenlange observatie trok het panel van experts haar diagnostische conclusies. De consensus met eerdere diagnoses was armoeiig: slechts drie van de vijf psychiatrische patiënten, maar ook twee van de gezonde proefpersonen kregen een diagnose toebedeeld. Hoe kleinschalig de aanpak van de documentairemakers ook was, hun project illustreert dat ongerichte observatie psychologen voor een formidabel demarcatieprobleem stelt: welk gedrag valt nog binnen de bandbreedte van het normale en wanneer is sprake van abnormaal gedrag? Wie lang genoeg kijkt, zal onherroepelijk iets afwijkends zien.

Een fraaie demonstratie daarvan is te vinden in de beschouwing die de Groningse hoogleraar Douwe Draaisma (2016) wijdt aan Ted Kazynski, de man die bekend staat als de *unabomber*, omdat hij een reeks van bomaanslagen pleegde op slachtoffers die werkzaam waren aan universiteiten. Hele pelotons van psychologen en psychiaters bogen zich over welke diagnose Kazynski ten deel moest vallen. Ze deden dat op grond van uitgebreide dossiers en sommigen waren in de gelegenheid om urenlang met Kazynski te spreken en hem te observeren. Zijn buurtgenoten vonden Kazynski een normale, zij het excentrieke man en zijn moeder bezwoer dat hij een gelukkige jeugd had gehad. Maar de psychologen en psychiaters meenden op geleide van hun observaties toch dat

Kazynski aan een stoornis leed, ofschoon hun diagnoses sterk uiteenliepen: van schizofrenie en autisme tot en met narcisme en psychopathie. Gebrekkige consensus, kortom. Elke expert kon zijn favoriete diagnose wel staven met saillante feiten over Kazynski's handel en wandel – zijn eenzijdige interesses, zijn slecht opgeruimde woonruimte, zijn talent voor wiskunde, zijn moeite om met krenkingen om te gaan. Die geïsoleerde feiten kregen pas reliëf door de diagnose en niet andersom.

Het probleem dat zich hier voordoet is dat van de *'overreliance on salient data'* (Wedding & Faust, 1989). Een ongemakkelijk gesprek met een psycholoog staat er garant voor dat zo'n psycholoog op enig moment per toeval het ene of andere saillante feit zal gaan observeren. Zo'n toevalsbevinding draagt niet bij aan de consensus. Daarom: de waarde van observatie is gering, tenzij de psycholoog op voorhand duidelijke hypothesen heeft geformuleerd en vastomlijnde ideeën heeft over welke observatiegegevens welke hypothesen ondersteunen dan wel tegenspreken. Niet ongerichte observatie, maar hypothese-toetsing is het devies.

## GESTRUCTUREERDE INTERVIEWS

Een alternatief voor ongerichte observatie is een gestructureerd interview met de onderzochte. Tijdens zo'n gestructureerd interview loopt de psycholoog een lijst van criteria – bijvoorbeeld die uit de *Diagnostic and Statistical Manual of Mental Disorders (DSM)* – langs om te bepalen of de onderzochte voldoet aan een diagnose. De *Structured Clinical Interview for DSM-5 (SCID)* is een voorbeeld. Veel clinici geven hoog op van instrumenten zoals de SCID. Daar is alle reden toe, want gestructureerde interviews bevorderen de consistentie en consensus (zie voor voorbeelden Mokros et al., 2018). Dat tweetal definieert, zoals gezegd, de ondergrens van fidele diagnostiek.

Toch is er ruimte voor verbetering. Terwijl voor sommige diagnoses – schizofrenie bijvoorbeeld – met gestructureerde interviews een hoge mate van consensus is te behalen, geldt dat voor andere stoornissen – depressie, gegeneraliseerde angststoornis, antisociale persoonlijkheidsstoornis – minder (Freedman et al., 2013; Vanheule et al., 2014; Chmielewski et al., 2015). Waar komt die variabiliteit vandaan? Een factor die telkens weer opduikt in onderzoek naar hoe psychologen tot diagnoses komen, is dat ze er allerlei stereotype ideeën op nahouden over de causale samenhang tussen symptomen (Garb, 2005; Flores, Cobos & Hagmayer, 2018). Het leidt ertoe dat ze aan sommige symptomen veel meer gewicht toekennen dan aan andere en soms ook symptomen menen te

## OF MISSCHIEN OP NAAR 2030?

Het overzicht van Harald Merckelbach en Brechje Dandachi-FitzGerald van valkuilen voor de diagnosticus en remedies daarvoor, kan gelden als verplichte literatuur voor bijvoorbeeld de Basisaantekening Psychodiagnostiek. Trouwens ook medisch-diagnostisch onderwijs, waar docenten en studenten het moeten stellen zonder psychometrische en cognitief-psychologische achtergrond, zou er om die reden alleen maar des te meer baat bij hebben. Merckelbach en Dandachi-FitzGerald leveren een staaltje metadiagnostiek, dus met de diagnosticus als cliënt. En dat niet alleen: de diagnostiek gaat hier vloeiend over in therapie, althans voor wie ontvankelijk is voor de uitdagingen van de schrijvers.

Boven hun betoog zweeft een discrepantie tussen twee paradigma's, de klassieke diagnostiek en de diagnostiek-cum-therapie zoals door Merckelbach en Dandachi-FitzGerald gepraktiseerd. Als voorwerp van hun betoog fungeert in hoofdzaak het klassiek diagnostisch paradigma. Daarin levert de diagnosticus een indicatiestelling voor een behandeling, passend bij een ziektebeeld. De opdracht voor de diagnosticus daarbij is waar mogelijk *false positives* en *false negatives*, voortkomend uit ongerichte en gerichte fouten te vermijden. De diagnose is geadresseerd aan de behandelaar (en diens opdrachtgever, zoals de verzekeraar); de patiënt is idealiter louter object, in de praktijk een potentiële stoorzender voor zover de diagnosticus afhankelijk is van informatie uit die bron (en de behandelaar afhankelijk is van *compliance*).

Nou, zo gaan Merckelbach en Dandachi-FitzGerald hier zelf dus niet te werk. Ze rapporteren rechtstreeks aan de patiënt, in casu de diagnosticus. Is dat paradigma misschien breder toepasbaar? Merckelbach en Dandachi-FitzGerald bespreken wel zoiets, al is dat een beetje bij wijze

van *afterthought*, in de paragraaf getiteld 'therapeutische diagnostiek'. Ze hebben er hun reserves over. Ze vermelden onderzoek waaruit blijkt dat eventuele misinformatie niet onschuldig is maar averechts werkt. Maar daarbij hanteren ze in feite een geradicaliseerd scenario, waarin het de cliënt is die de rol van de professional zou moeten overnemen -- en aan die opdracht niet kan voldoen. Merk op dat dit niet het scenario is dat ze in hun benadering van ons als lezer volgen; ik althans kan daar geen poging tot verspreiden van *fake news* in ontdekken, op de manier waarop dat in hun experimenten wel het geval was.

Wel aanbevelenswaardig is een scenario waarin de diagnosticus de cliënt of patiënt als medeverantwoordelijk gesprekspartner bij het proces betreft. Daarbij blijven de klassieke eisen van vermijden van valkuilen, die aan *learned professional* moeten worden gesteld, onverkort gehandhaafd. De cliënt is gesprekspartner (in plaats van alleen object), maar zelf geen professional, zelfs niet wanneer het een collega is.

Zo komen we misschien toch een stapje verder. Zelfs in medische kringen daagt het besef dat de patiënt zelf bijvoorbeeld moet worden betrokken bij de afweging van behandelingen, met inbegrip van ongewenste neveneffecten. Immers, die taxatie van subjectief verwacht nut is ter competentie van niemand anders dan dat subject. In beschouwingen over psychotherapie, en vermoedelijk nog meer in de dagelijkse praktijk, tekent zich een ontwikkeling af waarin de cliënt een persoon is, in de geest van de psychologische beroepscode. Op naar 2030.

*Wim Hofstee is emeritus-hoogleraar psychologie aan de Rijksuniversiteit Groningen. E-mail: w.k.b.hofstee@rug.nl.*

hebben waargenomen die er simpelweg niet waren.<sup>6</sup> Het is aannemelijk dat deze subjectieve weging afname en interpretatie van gestructureerde interviews stuurt – iets dat vooral manifest zal worden als vakgenoten uiteenlopende stereotypen in hun achterhoofd hebben (Garb, 2005). Een manier om daar iets aan te doen is door aan klinici rijke

beschrijvingen van prototypische diagnoses ter hand te stellen en hen te instrueren de overeenkomst tussen de onderzochte en de prototypes op een glijdende schaal te scoren (DeFife et al., 2013).

Leunen op discutabele stereotypen kan bijdragen aan wat in de literatuur *base rate neglect* heet (Richards et al., 2015). Neem dit realistische geval. Een psycholoog buigt zich op verzoek van de rechter-commissaris over het geval van een moeder die verzuimde haar baby uit het kinderbadje te

6 Zie voor een experimentele demonstratie hiervan: Weine en Kim (2018).



halen, waardoor de baby verdrong. De vrouw is dood door schuld ten laste gelegd, maar heeft ze misschien een stoornis waardoor ze haar baby vergat? De moeder vertelt dat zij aan migraine lijdt, dat er epilepsie in haar familie voorkomt, dat ze onlangs haar baan verloor en nu met financiële problemen worstelt. Het rapport van een neuroloog meldt dat de moeder veel functionele klachten heeft, waaronder pseudo-epilepsie. Tegen de tijd dat deze informatie de psycholoog heeft bereikt, is er een gerede kans dat die denkt in de richting van dissociatieve amnesie of *transient global amnesia*. Dat soort stereotypen zal de inhoud van het gestructureerde interview bepalen en het is denkbaar dat daarbij alle vragen handelend over mogelijke verslavingen worden overgeslagen. Dat zou een vorm van *base rate neglect* zijn, omdat de a priori kans dat de vrouw een alcoholprobleem heeft (7-9% van de populatie) groter is dan de kans dat zij lijdt aan dissociatieve amnesie (< 3%) of *transient global amnesia* (< 0.001%). Kennis nemen van lijsten met prevalentiecijfers voor de verschillende aandoeningen, symptomen en stoornissen is de remedie bij uitsteking tegen *base rate neglect* (Richards et al., 2015; Tiemens et al., 2018).

## Als er iets flonkert in de etalage van de psychologie, zijn het wel haar tests

### TESTS

Als er iets flonkert in de etalage van de psychologie, zijn het wel haar tests, want daarmee kan een hoge graad van consistentie en consensus worden bereikt.<sup>7</sup> Het overzichtsartikel van Meyer en collega's (2001) gaat een stap verder. Deze auteurs betrokken 125 meta-analytische studies naar de diagnostische slagkracht van de meest uiteenlopende tests – van de WAIS tot de MMPI – in hun beschouwing. De conclusie: 'Psychological test validity is comparable to medical test validity' (Meyer et al., 2001, p. 128) – een aanbeveling derhalve

om psychodiagnostiek te enten op tests. Het lijkt ons een verstandig uitgangspunt, maar wel met de kanttekening dat aan het gebruik ervan allerlei problemen kunnen kleven. Een daarvan doet zich gelden als de onderzochte op een test ver beneden de norm scoort en aan die bevinding een grote betekenis wordt toegekend. Het zwakke punt hier is dat uitschieters naar beneden heel wel nietszeggende fluctuaties kunnen zijn.

De Amerikaanse neuropsycholoog Binder en zijn collega's (2009) geven een overzicht van diverse studies waarin grote groepen volstrekt normale proefpersonen aan testbatterijen werden onderworpen. De gebruikelijke uitkomst was dat enkele tientallen procenten van hen ondermaats presteerde op tenminste een van de testonderdelen. Wie aan zulke toevalsbevindingen diagnostische waarde verbindt, kan op een dwaalspoor geraken. Dat valt te voorkomen door scores op de verschillende tests in hun onderlinge samenhang te bezien: het gaat niet om incidentele uitbijters, maar om het profiel. Het onlangs door Nederlandse neuropsychologen ontwikkelde en voor alle psychologen beschikbaar gestelde online-programma ANDI (*Advanced Neuropsychological Diagnostic Infrastructure*) biedt een probaat middel tegen overinterpretatie van geïsoleerde testcores.<sup>8</sup> Het idee achter ANDI is dat de diagnostische combinaties van testcores *upload*; het programma zet met multivariate statistiek het scoreprofiel van de onderzochte af tegen normgegevens en beziet langs deze weg of het profiel afwijkend is. ANDI is een prachtig initiatief om consistentie en consensus te vergroten (Agelink van Rentergem, 2018).

Een ander probleem dat zich bij tests kan voordoen is hun gevoeligheid voor *response bias*. Jazeker, de psycholoog kan aan de onderzochte een IQ-test, geheugentaken en allerlei andere cognitieve tests en zelfrapportage-vragenlijsten geven. Maar de scores van de onderzochte daarop laten zich pas interpreteren als de psycholoog deze preliminaire vraag onder ogen ziet: heeft de onderzochte zijn best gedaan om eerlijk te zijn? Onderzochten kunnen om allerlei redenen daarmee de hand lichten. Het kan zijn dat de onderzochte leerling geen zin heeft en de tests lukraak invult (*careless responding*; Meyer et al., 2013). Of de verdachte dikt zijn klachten aan of zuigt ze zelfs uit de duim (*faking bad*; Merckelbach, 2015). Of de sollicitant wil zo supernormaal mogelijk overkomen en ontkent allerlei minder fraaie eigenschappen (*faking good*; Griffith, Chmielowski & Yoshita, 2007).

7 Bij wijze van voorbeeld: Parker, Hanson & Hunsley (1988) middelden geschatte coëfficiënten voor consistentie en consensus voor veel gebruikte tests (zoals de mmpi en wais) en vonden waarden van ruim boven de 0.80.

8 <http://www.andi.nl/home-page/>

Zulke vertekeningen doen zich op ruime schaal voor. Maar net zoals economen vaak geloven in de fictie van Econs – consumenten baseren hun beslissingen volledig op boekhoudkundige afwegingen – zweren psychologen nogal eens bij Psycons. Als je Psycons vraagt om een schaal in te vullen, dan zullen ze met grote toewijding introspectie bedrijven en vervolgens goudeerlijke en precieze antwoorden genereren, zonder acht te slaan op hun belangen en die van de vraagsteller. Psycons bestaan natuurlijk niet en de ijzere consequentie hiervan is dat elke vorm van psychodiagnostiek om te beginnen moet kijken naar of testscores door een slordige testattitude, veinzen of mooipraterij worden getroebleerd. Die stap krijgt een ronduit stiefmoederlijke bedeling in handboeken zoals *Psychologische diagnostiek in de gezondheidszorg* (Luteijn & Barelds, 2018).<sup>9</sup>

Maar het goede nieuws is dat er ook in ons land de nodige tests beschikbaar zijn om die preliminaire stap wel te zetten. Het gaat dan om zogenaamde symtroomvaliditeitstests die de diagnosticus in de testbatterij kan – ja, moet – opnemen. Enkel als de onderzochte op deze tests niet afwijkend scoort, staat het licht op groen om over te gaan tot interpretatie van alle andere diagnostische gegevens. Het voert te ver om symtroomvaliditeitstests hier in detail te bespreken. We volstaan met de opmerking dat ze de potentie hebben om consistentie en consensus op te krikken.

De superieure introspectie van Psycons is lang niet iedereen gegeven. Erger nog, sommige onderzochten zullen al snel tegen de grenzen van hun zelfinzicht aanlopen en om die reden tests onnauwkeurig beantwoorden. Deze onnauwkeurigheid is van een andere orde dan gebrekkige symtroomvaliditeit; ze wijst op een metacognitief probleem. Een probaat middel om dit te repareren is door gebruik te maken van gegevens die door goed geïnformeerde derden over, bijvoorbeeld, het middelengebruik of de achterdochtigheid van de onderzochte worden aangedragen. Dat hun oordelen en die van de onderzochte uit de pas lopen, is niet ongewoon. Maar in plaats van zulke discrepanties terzijde te schuiven als ruis, zijn ze ook te gebruiken als psychodiagnostisch materiaal. Dat is wat het onderzoek van de Nederlandse psychologen Mosterman en Hendriks (2011) laat zien. Geïnspireerd door het pionierswerk van de Groningse hoogleraar Willem Hofstee (1994) vroegen ze patiënten met

psychische problemen om zichzelf te scoren op een lijst die dimensies van de *Big Five* meet. Informanten gaven op dezelfde lijst eveneens scores, maar dan voor de patiënten. Discrepanties tussen scores deden zich vooral voor bij introverte, hostile en depressieve patiënten met persoonlijkheidsproblemen.

Hofstee en collega's (2018) ontwikkelden inmiddels een persoonlijkheidsschaal die drie perspectieven omvat<sup>10</sup>: wat onderzochten over zichzelf beweren, wat anderen over de onderzochten zeggen en wat onderzochten denken dat anderen over hen menen te weten. Het zou ons niet verwonderen als dit instrument een psychometrisch juweeltje blijkt te zijn, dat ook therapeutische betekenis heeft. Mosterman en Hendriks (2011) kwamen in hun artikel al met de interessante suggestie om de discrepanties tussen wat de onderzochte over zichzelf beweert en wat informanten zeggen tot therapeutisch vehikel te maken. Zulke discrepanties kunnen duiden op een gebrekkige empathie en een onvermogen om het standpunt van anderen in te nemen, belangrijke thema's voor een therapeutisch gesprek. Wat ons brengt bij psychodiagnostiek als vorm van therapeutische consultatie.

## THERAPEUTISCHE DIAGNOSTIEK

Een hoge graad van consistentie en consensus nastreven – met alles wat daarbij hoort (zie tabel 1) – is een belangrijk kwaliteitskenmerk van psychodiagnostiek. Maar minstens zo belangrijk is wat psychodiagnostiek teweeg kan brengen. In duur jargon: haar consequentiële validiteit (Bornstein, 2017<sup>11</sup>). Interessant in dit verband is onderzoek naar therapeutische diagnostiek – *therapeutic assessment* – zoals beschreven door de Amerikaanse psycholoog Stephen Finn (2017). Dat is een vorm van diagnostiek waarbij de psycholoog in samenspraak met de onderzochte de belangrijke onderzoeksvragen formuleert, vervolgens psychodiagnostisch onderzoek verricht en tenslotte de resultaten daarvan uitvoerig bespreekt met de onderzochte. Het doel van die laatste stap is om de onderzochte inzicht te geven in aanknopingspunten voor veranderingen ten positieve. Uit de meta-analyse van Poston en Hanson (2010; maar zie ook Lilienfeld, Garb & Wood, 2011) blijkt dat zo'n aanpak inderdaad zoden aan de

9 Hoe anders is dat bijvoorbeeld in het Verenigd Koninkrijk, waar de British Psychological Society het belang van symtroomvaliditeitstests benadrukt in een beleidsdocument: British Psychological Society (2009). *Assessment of effort in clinical testing of cognitive functioning in adults.*

10 Zie voor deze *Drie Perspectieven Persoonlijkheden Questionnaire* (3PPQ): <https://3ppq.nl>.

11 Volledigheidshalve merken we op dat de eerder genoemde Willem Hofstee (2009) zich al veel eerder en uitvoeriger heeft uitgelaten over consequentiële validiteit.



## Het idee van therapeutische diagnostiek is sympathiek, maar wat als de diagnosticus kwestieuze tests gebruikt en de onderzochte voorziet van misleidende feedback?

dijk zet in termen van symptoomafname, verminderd middelengebruik en allerlei andere parameters: onderzochten profiteren er in aanzienlijke mate van (Cohen's  $d = 0.42^{12}$ ).

Er is wel geopperd om therapeutische diagnostiek halverwege de behandeling van patiënten te positioneren om zo vaart aan die behandeling te geven. Voor deze gedachte bestaat enige steun (Smith et al., 2015), maar meer onderzoek naar hoe psychodiagnostiek en behandeling effectief met elkaar zijn te verknopen is nodig. Dat geldt evenzeer voor de potentiële schaduwkant ervan. Dan hebben we het over het risico van diagnostische misinformatie. Want het idee van therapeutische diagnostiek is sympathiek, maar wat als de diagnosticus kwestieuze tests gebruikt en de onderzochte voorziet van misleidende feedback? Uit eigen onderzoek weten we dat mensen misleidende feedback over hun symptomen vrij makkelijk accepteren (Merckelbach et al., 2018). In dat onderzoek lieten we gezonde proefpersonen een test invullen met vragen over allerhande symptomen zoals concentratiestoornissen. Daarna gaven we hen tijdens een interview pertinent onjuiste feedback. We vroegen bijvoorbeeld: 'Kunt u meer vertellen over de concentratieproblemen die u heeft?', terwijl de proefpersonen eerder hadden ontkend die te hebben. Veel proefpersonen (87%) accepteerden deze misinformatie. Een aantal van hen (27%) confabuleerde ook oorzaken voor de aanvankelijk non-existente symptomen zoals: 'Ik heb de laatste tijd teveel dingen aan mijn hoofd.' Ook niet onbelangrijk: na een aantal dagen hadden deze proefpersonen de misinformatie geïnternaliseerd, wat bleek uit hogere scores op de betreffende symptomen.

Bevatelijkheid voor diagnostische misinformatie komt niet zo zeer voort uit goedgelovigheid, maar eerder uit ambiguïteit. Mensen vinden het lastig om de intensiteit en frequentie van hun interne ervaringen – waaronder symptomen – goed in te schatten en zijn bereid om daarvoor hun licht op te steken bij anderen. Dat geldt zelfs als die

anderen minimaal aanwezig zijn. In een onlangs uitgevoerd experiment (Van Helvoort et al., in voorbereiding) lieten we bijvoorbeeld proefpersonen symptoomlijsten invullen, waarna een computer algoritme hen zogenaamd daarover feedback zou geven. We vertelden de proefpersonen dat het algoritme volop in ontwikkeling was en verre van perfect. Voor sommige symptomen ('controls') voorzag het algoritme proefpersonen van correcte informatie, maar voor andere symptomen ('targets') van escalerende feedback. Meer dan de helft van de proefpersonen (66%) accepteerde de misleidende feedback en bij een her-test scoorden deze proefpersonen ook hoger op specifiek de symptomen waarover ze misleidende informatie hadden ontvangen (zie figuur 1).

We benadrukken het onderzoek naar de effecten van misleidende feedback omdat psychologen en leken steeds makkelijker toegang krijgen tot elektronische zelftests die snelle en goedkope diagnostiek in het vooruitzicht stellen. De kwaliteit daarvan is een punt van zorg. Illustratief zijn zogenaamde *sleep trackers*, makkelijk te downloaden apps die pretenderen de slaapkwaliteit te meten. Er zijn inmiddels in de literatuur gevallen beschreven van mensen wier slaapproblemen volkomen uit de hand liepen onder invloed van de dubieuze feedback van deze *sleep trackers* (Baron et al., 2017). Dat is een vorm van consequentiële validiteit die je liever niet wilt.

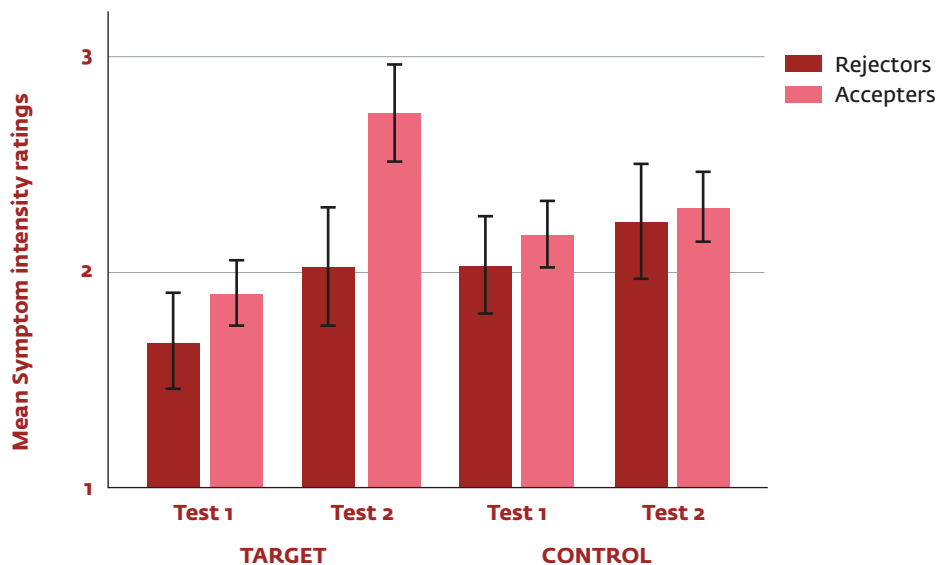
### BESLUIT

Psychodiagnostiek behoort tot het beste wat de psychologie te bieden heeft. Ze kan een zinnig antwoord geven op belangrijke vragen die mensen over zichzelf en anderen stellen. Het vereist wel dat psychodiagnostiek ligt ingebed in een protocol, observaties een hypothese-toetsende rol spelen, verstandig gebruik wordt gemaakt van gestructureerde interviews en tests, en – indien mogelijk – informatie van derden wordt verzameld. Zo vormgegeven, stelt psychodiagnostiek hoge eisen aan de deskundigheid van psychologen. Onderwijzen we onze psychologiestudenten daarin voldoende?

Bornstein (2017) wijst op de afkalvende positie van psychodiagnostiek in de curricula van Amerikaanse psycho-

12 Deze effectgrootte is volgens de gebruikelijke standaarden gemiddeld, maar daar staat tegenover dat de interventie uit niet meer dan drie sessies bestond.

FIGUUR 1.



Proefpersonen (N = 80) vulden een symptoomlijst in (Test 1) en kregen vervolgens escalerende feedback over sommige symptomen (Target) en correcte feedback over andere symptomen (Controls). 49 proefpersonen (66%) accepteerden de misleidende feedback (*accepters*); de anderen spraken de misleidende feedback tegen (*rejecters*). Na een intermezzo vulden de proefpersonen de symptoomlijst andermaal in (Test 2). Symptomen werden gescoord op een schaal van 1 (nooit last van) tot 5 (voortdurend last van). De figuur toont gemiddelde scores en daarbij behorende 95%-betrouwbaarheidsintervallen.

logie-opleidingen en onze indruk is dat bij ons de vlag er niet veel beter bij hangt. Daar staat de opmars van psychodiagnostische piraterij tegenover: televisieprogramma's waarin op basis van dubieuze tests huwelijkspartners aan elkaar worden gekoppeld (*Married at first sight*<sup>13</sup>), bedrijfjes (*Braincompass*<sup>14</sup>) die met behulp van DNA-analyse geschikte *salesmanagers* rekruteren en medische faculteiten die met ondoorzichtige procedures kandidaten voor de geneeskunde-studie<sup>15</sup> selecteren – om enkele willekeurige voorbeelden te

noemen. Tegen dat feit te hoop lopen vereist een soort beginselvastheid die we niet van de doorsnee onderwijs-directeur aan een psychologiefaculteit verwachten. Maar misschien toch wel het besef dat psychodiagnostiek meer is dan wat zelfrapportagetests afnemen en scoren. Om daaraan bij te dragen schreven we dit artikel.

#### OVER DE AUTEURS

Harald Merckelbach is hoogleraar rechtspsychologie en verbonden aan de Faculteit Psychologie en Neurowetenschappen van Universiteit Maastricht. Brechje Dandachi-FitzGerald is klinisch psycholoog en universitair docent, verbonden aan de Faculteit Psychologie en Neurowetenschappen van Universiteit Maastricht. Correspondentie aangaande dit artikel: [h.merckelbach@maastrichtuniversity.nl](mailto:h.merckelbach@maastrichtuniversity.nl).

13 Zie daarover Vittorio Busato: 'Vulgariserende wetenschap.' *Metro* 11 april 2018.

14 Zie daarover Anne Vegterlo: 'Zit de baan in je genen?' *NRC Handelsblad* 2 november 2014. Het motto van *Braincompass* is trouwens: 'inzicht in jouw biologische identiteit begint met inzicht in jouw DNA'.

15 Zie daarover Harald Merckelbach: 'Wie bluft mag dokter worden.' *NRC Handelsblad* 28 februari 2015.

# Literatuur

- Agelink van Rentergem, J. (2018). *Statistical advances in neuropsychology*. Amsterdam, UvA proefschrift.
- Baca-Garcia, E., Perez-Rodriguez, M. M., Basurte-Villamor, I., Del Moral, A. L. F., Jimenez-Arriero, M. A. et al. (2007). Diagnostic stability of psychiatric disorders in clinical practice. *British Journal of Psychiatry*, 190, 210-216.
- Baron, K. G., Abbott, S., Jao, N., Manalo, N. & Mullen, R. (2017). Orthosomnia: Are some patients taking the quantified self too far? *Journal of Clinical Sleep Medicine*, 13, 351-354.
- Binder, L. M., Iverson, G. L. & Brooks, B. L. (2009). To err is human: "Abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, 24, 31-26.
- Block, A. R., Ohmmeiss, D. D., Guyer, R. D., Rashbaum, R. F. & Hochschuler, S. H. (2001). The use of presurgical psychological screening to predict the outcome of spine surgery. *The Spine Journal*, 1, 274-282.
- Bornstein, R. F. (2017). Evidence-based psychological assessment. *Journal of Personality Assessment*, 99, 435-445.
- Chmielewski, M., Clark, L. A., Bagby, R. M. & Watson, D. (2015). Method matters: Understanding diagnostic reliability in DSM-IV and DSM-5. *Journal of Abnormal Psychology*, 124, 764-769.
- DeFife, J. A., Peart, J., Bradley, B., Ressler, K., Drill, R. & Westen, D. (2013). Validity of prototype diagnosis for mood and anxiety disorders. *JAMA Psychiatry*, 70, 140-148.
- Draaisma, D. (2016). *Als mijn geheugen me niet bedriegt*. Groningen: Historische Uitgeverij.
- Eurelings, L. S. M. & Van Os, J. (2018). Medicalisering en ondoelmatige zorg in de GGZ: De rol van stoornisgericht denken en vergoeden. *Nederlands Tijdschrift voor Geneeskunde*, 162, D 3534.
- Finn, S. E. (2007). *In our clients' shoes, theory and techniques of therapeutic assessment*. New York: Routledge.
- Flores, A., Cobos, P. L. & Hagmayer, Y. (2018). The diagnosis of mental disorders is influenced by automatic causal reasoning. *Clinical Psychological Science*, 6, 177-188.
- Freedman, R., Lewis, D. A., Michels, R., Pine, D. S., Schultz, S. K. et al. (2013). The initial field trials of DSM-5: new blooms and old thorns. *American Journal of Psychiatry*, 170, 1-5.
- Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology*, 1, 67-89.
- Griffith, R. L., Chmielewski, T. & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, 36, 341-355.
- Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality*, 8, 149-162.
- Hofstee, W. K. B. (2009). Restyling personality assessments. In L. B. Pincus & M. V. Lopez (Eds), *Personality assessment: New research* (pp. 223-235). Hauppauge, NY: Nova Science Publishers.
- Hofstee, W. K. B., Mosterman, I., Hendriks, J., Kiers, H. & Brokken, F. (2018). Persoonlijkheid in drie perspectieven: Een instrument. *De Psycholoog*, 53, 50-51.
- Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E. & Regier, D. A. (2012). DSM-5: how reliable is reliable enough? *American Journal of Psychiatry*, 169, 13-15.
- Lilienfeld, S. O., Garb, H. N. & Wood, J. M. (2011). Unresolved questions concerning the effectiveness of psychological assessment as a therapeutic intervention: Comment on Poston and Hanson (2010). *Psychological Assessment*, 23, 1047-1055.
- Luteijn, F. & Barelds, D. (2018). *Psychologische diagnostiek in de gezondheidszorg*. Hoofddorp: Boom.
- Merckelbach, H. (2015). Vijf mythes over simulanten: En waarom we ze geloven. *De Psycholoog*, 50, 11-19.
- Merckelbach, H., Dalsklev, M., van Helvoort, D., Boskovic, I. & Otgaar, H. (2018). Symptom self-reports are susceptible to misinformation. *Psychology of Consciousness: Theory, Research, and Practice*, 5, 384-397.
- Meyer, J. F., Faust, K. A., Faust, D., Baker, A. M. & Cook, N. E. (2013). Careless and random responding on clinical and research measures in the addictions: A concerning problem and investigation of their detection. *International Journal of Mental Health and Addiction*, 11, 292-306.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L. et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128-165.
- Mokros, A., Habermeyer, E. & Küchenhoff, H. (2018). The uncertainty of psychological and psychiatric diagnoses. *Psychological Assessment*, 30, 556.
- Mosterman, R. M. & Hendriks, A. J. (2011). Self-other disagreement in personality assessment: Significance and prognostic value. *Clinical Psychology & Psychotherapy*, 18, 159-171.
- Parker, K. C., Hanson, R. K. & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin*, 103, 367.
- Poston, J. M. & Hanson, W. E. (2010). Meta-analysis of psychological assessment as a therapeutic intervention. *Psychological Assessment*, 22, 203-212.
- Proglora, Y. (2009). Mental illness and social stigma: notes on "How Mad Are You?" *Journal of Research in Medical Sciences*, 14, 331-334.
- Richards, P. M., Geiger, J. A. & Tussey, C. M. (2015). The dirty dozen: 12 sources of bias in forensic neuropsychology with ways to mitigate. *Psychological Injury and Law*, 8, 265-280.
- Schilders, M. R. & Ogloff, J. R. (2017). Stability of life-time psychiatric diagnoses among offenders in community and prison settings. *Journal of Forensic Psychiatry & Psychology*, 28, 133-154.
- Scott, K. & Lewis, C. C. (2015). Using measurement-based care to enhance any treatment. *Cognitive and Behavioral Practice*, 22, 49-59.
- Smith, J. D., Eichler, W. C., Norman, K. R. & Smith, S. R. (2015). The effectiveness of collaborative/therapeutic assessment for psychotherapy consultation: a pragmatic replicated single-case study. *Journal of Personality Assessment*, 97, 261-270.
- Spaanjaars, N. L., Groenier, M., van de Ven, M. O. & Witteman, C. L. (2015). Experience and diagnostic anchors in referral letters. *European Journal of Psychological Assessment*, 31, 280-286.
- Thomas, R. P. & Lawrence, A. (2018). Assessment of expert performance compared across professional domains. *Journal of Applied Research in Memory and Cognition*, 7, 167-176.
- Tiemens, B., Wagenvoorde, R. & Witteman, C. (2018). Waarom iedere psycholoog de regel van Bayes moet kennen. *GZ-psychologie*, 3, 17-23.
- Timmermans, M. (2018). Herstel. *De Psychiater*, augustus, 10-12.
- Van Helvoort, D., Otgaar, H. & Merckelbach, H. (in voorbereiding). Worsening of self-reported symptoms through suggestive feedback.
- Vanheule, S., Desmet, M., Meganck, R., Inselegers, R., Willemsen, J. et al. (2014). Reliability in psychiatric diagnosis with the DSM: old wine in new barrels. *Psychotherapy and Psychosomatics*, 83, 313-314.
- von dem Knesebeck, O., Bönnte, M., Siegrist, J., Marceau, L., Link, C. & McKinlay, J. (2010). Diagnose und Therapie einer Depression im höheren Lebensalter-Einflüsse von Patienten- und Arztmerkmalen. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 60, 98-103.
- Wedding, D. & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology*, 4, 233-265.
- Weine, E. R. & Kim, N. S. (in druk). Systematic distortions in clinicians' memories for client cases: Increasing causal coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.