

Redactioneel

De Commissie Testaangelegenheden Nederland (COTAN) over validiteitstests: wat kan de rechter ermee?

18

De Commissie Testaangelegenheden Nederland (COTAN), ressorterend onder het Nederlands Instituut van Psychologen, beoordeelt de kwaliteit van psychologische tests.¹ Daarmee voorziet deze commissie in een behoefte. Gebruikers van deze tests – maar ook de mensen die de tests ondergaan – willen immers weten of ze solide genoeg zijn om er conclusies aan te verbinden. Maar voor zogeheten validiteitstests² is een oordeel van de COTAN vaak niet beschikbaar, verouderd of beperkt onderbouwd. Toch maken psychologen gebruik van deze categorie van tests, vooral wanneer zij als deskundigen psychodiagnostisch onderzoek verrichten in het kader van civiele of strafrechtelijke procedures. In zo'n geval kan een ongefundeerd of afwezig COTAN-oordeel nadelig uitpakken voor de psycholoog. Waarom dat een slechte zaak is, leggen we hieronder uit. Dat vergt allereerst een korte toelichting op wat validiteitstests zijn.

1. Ruis

Als psycholoog kun je testen tot je een ons weegt, maar als testuitkomsten vol ruis zitten, is hun informatiegehalte nihil. Ruis kan allerlei vormen aannemen: de onderzochte kan tests lukraak invullen, symptomen aandikken of zich onvoldoende inspannen om naar beste kunnen te presteren op neuropsychologische taken. In al die situaties kunnen testresultaten maar beter niet worden meegenomen in het diagnostisch oordeel over een verdachte of claimant. Net zomin kunnen ze bijdragen aan de behandeladviezen voor de patiënt of de delinquent. De psycholoog zou over voldoende kennis moeten beschikken om te weten dat iemand die de testuitkomsten naar zijn hand probeert te zetten, niet direct in de kijker hoeft te lopen. Zo iemand kan een authentieke en coöperatieve indruk maken en evenzogoed symptomen fabuleren. Het omgekeerde komt trouwens ook voor: de psycholoog kan bijvoorbeeld twijfelen aan de bereidheid van de onderzochte om zich naar behoren in te spannen en dan kan toch blij-

ken dat de onderzochte naar vermogen heeft gepresteerd.³

Of testuitkomsten door ruis vertroebeld zijn, laat zich vaststellen met zogenaamde validiteitstests. De vakliteratuur over dit type test heeft in de afgelopen twintig jaar een stormachtige groei doorgemaakt. Naar schatting een kwart van alle artikelen in neuropsychologische vakbladen heeft betrekking op de ontwikkeling, ijking en kalibratie van validiteitstests, waarvan er inmiddels een dikke vijftig in psychometrisch opzicht redelijk goed zijn gedocumenteerd.⁴ De opmars van validiteitstests is overigens niet alleen zichtbaar in de neuropsychologie, maar tekent zich ook af in de forensische psychologie.⁵

2. Overrapporteren

Validiteitstests bestaan er in soorten en maten. Om te beginnen zijn er tests die peilen of door de onderzochte symptomen worden gerapporteerd in een mate die onwaarschijnlijk is. De ook in ons land regelmatig

* Prof. dr. H.L.G.J. Merckelbach is hoogleraar rechtspsychologie aan de Universiteit Maastricht en tevens redacteur van dit blad.

Prof. dr. B. Dandachi-FitzGerald is als bijzonder hoogleraar Psychotherapie verbonden aan de Open Universiteit Heerlen. Zij is tevens universitair hoofddocent aan de Universiteit Maastricht, alsook hoofdopleider Psychotherapie bij RINO Zuid, Eindhoven. De auteurs danken dr. S. Roza voor haar commentaar op een eerdere versie van dit stuk.

1. cotandocumentatie.nl/beoordelingen/.

2. Zie daarvoor: H. Merckelbach & B. Dandachi-FitzGerald, 'Symptoomvaliditeitstests in de rechtszaal', *EeR* 2021, afl. 3, p. 105-111. En voorts: H. Merckelbach, 'Symptoomvaliditeitstests in arbeidsongeschiktheidszaken', *EeR* 2023, afl. 6, p. 183-186.

3. B. Dandachi-FitzGerald e.a., 'Neuropsychologists' ability to predict distorted symptom presentation', *Journal of Clinical and Experimental Neuropsychology* 2017, 39(3), p. 257-264.

4. P.K. Martin e.a., 'A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection', *The Clinical Neuropsychologist* 2020, 34(1), p. 88-119; Institute of Medicine, 'Psychological testing in the service of disability determination', National Academies Press 2015. Zie nap.nationalacademies.org/catalog/21704/psychological-testing-in-the-service-of-disability-determination.

5. T. Merten e.a., 'Symptom and performance validity assessment in European countries: An update', *Psychological Injury and Law* 2022, 15(2), p. 116-127.

gebruikte *Structured Inventory of Malingered Symptomatology* (SIMS) kan als voorbeeld dienen.⁶ De crux van deze tests is dat ze de onderzochte een reeks van zeldzame of niet bestaande symptomen voorhouden en dan nagaan of die op ruime schaal worden omarmd. Is dat het geval, dan bestaat er alle reden voor de psycholoog om kritisch stil te staan bij wat de onderzochte spontaan rapporteert over klachten, maar ook bij de symptomen die betrokkene aankruist op andere vragenlijsten. Dit type validiteitstest is zodoende van prioritair belang voor de psychodiagnostiek. Amerikaanse, maar ook Nederlandse beroepsverenigingen van neuropsychologen (Nederlands Instituut voor Psychologen, sectie Neuropsychologie, 2016⁷) prijzen het gebruik van deze tests aan, zeker – maar niet uitsluitend (zie onder) – als psychodiagnostiek plaatsvindt in de context van een deskundigenbericht. Zo'n aanbeveling is bovendien aan te treffen in de richtlijnen van het Nederlands Instituut voor Forensische Psychiatrie en Psychologie (NIFP).⁸ Maar hier ontstaat wel een probleem met het kwaliteitsoordeel van de COTAN. Die acht een test als de SIMS onder de maat. Er zouden, aldus de COTAN, geen goede normgegevens voor de SIMS bestaan en bovendien zou onduidelijk zijn of de test wel valide is, dat wil zeggen, meet wat hij wil meten.

3. Onderpresteren

Eenzelfde contradictie doet zich voor bij de tweede categorie van validiteitstests. Het gaat dan om cognitieve taken die onderpresteren meten. Een in dat verband vaak gebruikte taak is de *Amsterdamse Korte Termijn Geheugentest* (AKTG).⁹ De AKTG zal op leken de indruk wekken een lastige geheugentaak te zijn. Maar zelfs patiënten met een neurologische aandoening kunnen op deze taak een prima geheugenprestatie leveren. Van onderpresteren is sprake als iemand het op deze taak slechter doet dan wat je bij patiënten zou verwachten. In dat geval rijst de vraag in hoeverre de scores van de onderzochte op alle andere neuropsychologische

taken representatief zijn voor diens cognitieve capaciteiten. Validiteitstests als de AKTG kunnen zodoende informatief zijn, reden waarom de eerdergenoemde beroepsorganisaties ze dienstig vinden. Opnieuw denkt de COTAN daar in het geval van de AKTG beslist anders over, want haar oordeel luidt andermaal dat de kwaliteit van deze test te wensen overlaat: er zouden onvoldoende normgegevens voorhanden zijn en de betrouwbaarheid van deze test acht de COTAN suboptimaal.

4. Validiteitstests in 'de COTAN'

Validiteitstests zijn slecht vertegenwoordigd in de catalogus van testbeoordelingen die de COTAN aanlegt. In deze catalogus deelt de COTAN voldoende en onvoldoende uit aan zo'n 900 tests.¹⁰ Terwijl intelligentietests of depressievragenlijsten er ruimschoots in kunnen worden aangetroffen, zijn validiteitstests op de vingers van een hand te tellen. Dat is merkwaardig als bedacht wordt dat het gros van de meetinstrumenten voor intelligentie of depressie zelf geen ingebouwde validiteitsschaal bevat. Interpretatie van de gegevens die hiermee worden verkregen, veronderstelt zodoende dat ze accuraat zijn, dat wil zeggen niet vertroebeld door overrapporteren of onderpresteren. Het is naïef om aan te nemen dat zulke vertroebelingen uitsluitend voorkomen wanneer tests worden afgenomen in een forensische context.¹¹ De testgegevens van een niet-triviale minderheid van gezonde proefpersonen,¹² maar ook patiënten¹³ vertoont immers dit soort ruis. Dat maakt het allemaal des te vreemder dat de COTAN zo weinig plaats inruimt voor validiteitstests.

De wetenschappelijke literatuur biedt ondertussen meta-analyses¹⁴ voor de volgende validiteitstests, waarvan een aantal ook in ons land wordt gebruikt: de *Miller Forensic Assessment of Symptoms*¹⁵ (M-Fast), de *Personality Assessment Inventory* (PAI),¹⁶ de *Inventory of Problems-29* (IOP-29),¹⁷ de validiteitsschalen van

6. G.P. Smith & G.K. Burger, 'Detection of malingering: validation of the Structured Inventory of Malingered Symptomatology (SIMS)', *Journal of the American Academy of Psychiatry and the Law* 1997, 25(2), p. 183-189.
7. B. Dandachi-FitzGerald e.a., 'Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries', *Archives of Clinical Neuropsychology* 2013, 28(8), p. 771-783.
8. Nederlands Instituut voor Psychologen, sectie Neuropsychologie. (2016). *Richtlijn neuropsychologische expertise*. nip.nl/wp-content/uploads/2022/04/RichtlijnNPexpertise_definitief_didb.pdf.
9. R. Hoogerwerf e.a., *Best practice ambulant: Forensisch onderzoek en rapportage in het strafrecht voor volwassenen*, Nederlands Instituut voor Forensische Psychiatrie en Psychologie 2009.
10. B. Schmand e.a., *Amsterdamse Korte Termijn Geheugentest, AKTG. Handleiding. Opgavenboek*, Swets & Zeitlinger 1999.
11. Zie: cotandocumentatie.nl/beoordelingen/p/2/.
12. H. Merkelbach & B. Dandachi-FitzGerald, 'Psychodiagnostiek anno 2020', *De Psycholoog* 2020, 55(1), p. 34-44.
13. K.Y. An e.a., 'Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance?', *Archives of Clinical Neuropsychology* 2012, 27(8), p. 849-857.
14. J.J. Roor e.a., 'Performance validity test failure in the clinical population: A systematic review and meta-analysis of prevalence rates', *Neuropsychology Review* 2024, 34(1), p. 299-319.
15. Zie over de superieure bewijskracht van meta-analyses: R. Giard & H. Merkelbach, 'De ene deskundige is de andere niet: hoe de rechter empirisch gefundeerd bewijs kan waarderen', *NJB* 2018, afl. 3, p. 181-188.
16. D. Detullio e.a., 'A meta-analysis of the Miller Forensic Assessment of Symptoms Test (M-FAST)', *Psychological Assessment* 2019, 31(11), p. 1319-1328.
17. S.W. Hawes & M.T. Boccacini, 'Detection of overreporting of psychopathology on the Personality Assessment Inventory: a meta-analytic review', *Psychological Assessment* 2009, 21(1), p. 112-124; J.E. Kurtz & E.M. Pintarelli, 'The Daubert Standards for admissibility of evidence based on the Personality Assessment Inventory', *Psychological Injury and Law* 2024, 17(1), p. 105-116.
18. E. Puente-López e.a., 'The Inventory of Problems-29 (IOP-29): a systematic review and bivariate diagnostic test accuracy meta-analysis', *Psychological Assessment* 2023, 35(4), p. 339-352.

de MMPI,¹⁸ de *Advanced Clinical Solutions Word Choice Test*¹⁹ en de *Effort Index van de Repeatable Battery for the Assessment of Neuropsychological Status*.²⁰ Andere meta-analytische studies vergeleken de merites van verschillende validiteitstests.²¹ Daarnaast richtten sommige meta-analyses zich op populaties met een hogere kans op fout-positieve uitkomsten, zoals patiënten met psychotische symptomatologie.²² Weer andere meta-analyses schatten de prevalentie van afwijkende scores op validiteitstests in verschillende situaties, zoals de reguliere patiëntenzorg.²³ In de afgelopen jaren committeerden de studies in dit domein zich steeds vaker aan de uitgangspunten van *open science* door preregistratie in openbare databestanden zoals Prospero²⁴ en Open Science Framework (OSF).²⁵ In onderlinge samenhang bezien rechtvaardigt dit corpus aan studies de conclusie dat verschillende validiteitstests stevig verankerd zijn in de wetenschappelijke literatuur en psychometrisch behoorlijk presteren. Maar deze ontwikkeling is niet terug te vinden in de COTAN-lijst.

Daar waar de COTAN zich wél uitlaat over validiteitstests, verlaat ze zich soms op achterhaalde opvattingen. Zo meent de commissie dat de eerder genoemde SIMS bedoeld is om 'malingeren' vast te stellen. Dat was misschien de pretentie in de beginjaren, toen deze test nog uitsluitend bij delinquenten werd ingezet. Inmiddels zijn de meeste experts het er echter over eens dat de SIMS overrapportage van symptomen meet; de achterliggende oorzaken daarvan kunnen sterk variëren.²⁶

5. Oude onvoldoendes

In ons land worden de SIMS en de AKTG regelmatig gebruikt om vast te stellen of sprake is van vertroebelde testresultaten. Deze tests figureren, zoals gezegd, wél in de COTAN-lijst en krijgen daar ongunstige beoordelingen. Meer specifiek ontvangt de SIMS onvoldoendes voor begripsvaliditeit en normgegevens. De AKTG doet het volgens de COTAN nog beroerder, want deze test moet onvoldoendes incasseren voor begripsvaliditeit,

normgegevens en betrouwbaarheid. Wellicht waren al die onvoldoendes ooit, een decennium geleden, terecht. Inmiddels weerspiegelen ze niet langer de actuele stand van zaken in de wetenschappelijke literatuur. Niettemin blijven deze verouderde beoordelingen wel rouleren.

Zo komen de onvoldoendes van de COTAN voor de SIMS en de AKTG omstandig ter sprake in een artikel van twee rechtspsychologen over het (psychodiagnostisch) onderzoek dat psychiaters en gezondheidszorgpsychologen verrichten naar de geestesgesteldheid van Thijs H.²⁷ De rechtspsychologen besteden veel aandacht aan deze en andere gebruikte tests en de beoordeling daarvan door het COTAN en ze verwijten de forensische rapporteurs dat die daaraan voorbijgingen. In lijn met het gezag dat aan de COTAN wordt toegekend, lijken de auteurs stilzwijgend aan te nemen dat de SIMS en de AKTG nog steeds niet aan de maat zijn, want anders zou de COTAN haar oordeel wel hebben bijgesteld. Die aanname spoort evenwel niet met de vakliteratuur. Een uitvoerige bespreking daarvan valt buiten het bestek van dit redactioneel. Hier willen we bij wijze van voorbeeld slechts volstaan met de vaststelling dat ook in het geval van de SIMS er omvangrijke meta-analyses voorhanden zijn die een gunstiger – ofschoon zeker niet perfect – beeld laten zien.²⁸ Voor een aanzienlijk deel zijn deze meta-analyses overigens gebaseerd op Nederlands materiaal. Ons punt is dat de onvoldoendes van de COTAN hier obsoleet en daarom onbruikbaar zijn.

6. Wat betekent een onvoldoende?

Bij wijze van belangeloze bijdrage aan het maatschappelijk verkeer, steekt de COTAN de peilstok in psychologische tests en deelt dan voldoende en onvoldoendes uit. De elf faculteiten met een opleiding psychologie fourneren elk jaar ieder een paar duizend euro ten behoeve van dit werk. Zo'n blijf van vertrouwen wekt verwachtingen, maar we betwijfelen of ze ten volle worden waargemaakt. Als we het goed zien is meer dan de helft van alle COTAN-beoordelingen, die van de eerder

18. M. Aparcero e.a., 'Comparing the ability of MMPI-2 and MMPI-2-RF validity scales to detect feigning: a meta-analysis', *Assessment* 2023, 30(3), p. 744-760.
19. M.T. Bernstein e.a., 'A systematic review and meta-analysis of the diagnostic accuracy of the advanced clinical solutions word choice test as a performance validity test', *Neuropsychology Review* 2021, 31(2), p. 349-359.
20. R.D. Shura e.a., 'RBANS validity indices: a systematic review and meta-analysis', *Neuropsychology Review* 2018, 28(3), p. 269-284; W.F. Goette & H.E. Goette, 'A meta-analysis of the accuracy of embedded performance validity indicators from the repeatable battery for the assessment of neuropsychological status', *The Clinical Neuropsychologist* 2019, 33(6), p. 1044-1068.
21. Bijvoorbeeld C.D. Vickery e.a., 'Detection of inadequate effort on neuropsychological testing: a meta-analytic review of selected procedures', *Archives of Clinical Neuropsychology* 2001, 16(1), p. 45-73; M.J. Sollman & D.T. Berry, 'Detection of inadequate effort on neuropsychological testing: a meta-analytic update and extension', *Archives of Clinical Neuropsychology* 2011, 26(8), p. 774-789.
22. I. Ruiz e.a., 'A meta-analysis of neuropsychological effort test performance in psychotic disorders', *Neuropsychology Review* 2020, 30(1), p. 407-424.
23. Roor e.a. 2024.
24. Ruiz e.a. 2020; Roor e.a. 2024.
25. Bijvoorbeeld A. Robinson e.a., 'Racial and ethnic representation of U.S. Adults within PVT classification accuracy studies: A systematic review', december 2023; doi.org/10.17605/OSF.IO/KC8RF.
26. H. Merkelbach & B. Dandachi-FitzGerald, 'The (small) world of validity test research according to Leonhard and Leonhard (2024): a critique', *Psychological Injury and Law* 2024, 17, p. 439-449.
27. T. de Beuf & P. van Koppen, 'Het testen van Thijs: een gevalstudie van forensisch testgebruik', *De Psycholoog* 2024, 59(1), p. 14, Tabel 3. Zie ook B. Boer e.a., 'Een stukje van de puzzel', *De Psycholoog* 2025, 60(2), p. 34-39.
28. A. van Impelen e.a., 'The structured inventory of malingered symptomatology (SIMS): a systematic review and meta-analysis', *The Clinical Neuropsychologist* 2014, 28(8), p. 1336-1365. R.D. Shura e.a., 'Structured inventory of malingered symptomatology: a psychometric review', *Psychological Injury and Law* 2022, 15(1), p. 64-78.

genoemde validiteitstests inclusief, ouder dan tien jaar. Oude, wellicht achterhaalde, onvoldoendes prijzen niettemin op de voor velen makkelijk te raadplegen lijst van de COTAN. Wat het extra verwarrend maakt is dat in de nomenclatuur van de COTAN een onvoldoende twee nogal verschillende dingen kan betekenen. Het kan erop wijzen – en zo zullen de meeste gebruikers van de COTAN-lijst het verstaan – dat een test jammerlijk tekortschiet. Maar het kan ook betekenen dat de COTAN niet de hand heeft weten te leggen op relevante informatie: ‘onvoldoende’ staat dan voor ‘niet bij ons bekend’.

Op 7 juni 2024 vond er in Utrecht een consensusvergadering plaats van forensisch en klinisch psychologen over het gebruik van validiteitstests. Een van de vragen die aan de orde kwam was of de COTAN bruikbare informatie biedt als het om deze categorie van tests gaat. Uit het verslag van de vergadering: ‘De consensus is dat de COTAN geen goede kwaliteitstoets biedt [...]. De COTAN wordt – als het om dit terrein gaat – gezien als onvolledig, namelijk niet geïnformeerd door overzichtsartikelen aangaande het testen van symptoomvaliditeit. Ook de door de COTAN gehanteerde kwaliteitscriteria (bijvoorbeeld “normgegevens”) schieten [...] tekort [...]. Het ware goed om daarover met de COTAN in discussie te treden, want de geringe aandacht die COTAN aan validiteitstests schenkt, kan zich tegen de clinicus/forensisch expert keren zodra die besluit om zulke tests wel in te zetten.’²⁹ De experts stelden zich overigens unaniem op het standpunt dat het gebruik van validiteitstests volstrekt te rechtvaardigen valt, ook als er geen of slechts een verouderde COTAN-waardering voorhanden is, zolang vakpublicaties maar aantonen dat hun psychometrische eigenschappen deugen. Dat zal heel regelmatig het geval zijn. Hoe dan ook, het standpunt van deze experts over de COTAN contrasteert nogal met het beeld dat de COTAN van zichzelf heeft.³⁰

7. Geen KEMA

Gaandeweg heeft de COTAN een reputatie verworven die lijkt op die van het keuringsinstituut KEMA. Toen bijvoorbeeld een overheidsinstantie jaren geleden een aanbesteding uitschreef voor het leveren van psychologische tests gold als een van de inschrijvingscriteria: ‘De test(en) of testonderdelen dienen psychometrisch van goede kwaliteit te zijn. De kwaliteit wordt beoordeeld aan de hand van de criteria die de COTAN (Commissie Testaangelegenheden Nederland van het Nederlands In-

stituut van Psychologen/NIP) hanteert.’³¹ Ook in rechts-geschillen komt het wel voor dat eisers hun weigering om een test te ondergaan door een rechter gehonoreerd willen zien onder verwijzing naar het argument dat de test ‘niet-COTAN’ goedgekeurd is.³² Klaarblijkelijk laat de COTAN zich in zulke gevallen het gezag aanleunen om kwaliteitsoordelen over psychologische tests en taken te verstrekken.

Nu is het zeker prijzenswaardig als een commissie zich buigt over iets dat tot de kern van de psychologische beroepsuitoefening behoort. Maar wie kwaliteitsoordelen geeft, laadt verplichtingen op zich. Kwaliteitsoordelen dienen te worden gegeven door ter zake kundige experts, bij voorkeur na een hoor- en wederhoorprocedure. Want voorkomen moet worden dat kwaliteitsoordelen slecht onderbouwd zijn. Evenzeer moet worden voorkomen dat de *afwezigheid* van zulke oordelen wordt uitgelegd ten nadele van een test. We zijn er niet helemaal gerust op dat de werkwijze van de COTAN hieraan voldoet.³³

Een voorbeeld van wat er kan gebeuren, biedt het artikel van Schimmel en Van Koppen (2017).³⁴ Met de COTAN in de hand gaven deze auteurs forensisch psychologen een behoorlijke veeg uit de pan: omdat ze ten behoeve van hun pro justitia-rapportages tests hadden gebruikt die hetzij niet in de COTAN-lijst voorkomen, hetzij van de COTAN een onvoldoende hadden gekregen. Daarbij ging het onder meer om validiteitstests zoals de SIMS. De auteurs: ‘De conclusie is eenvoudig: in meer dan de helft van de gevallen wordt een test gebruikt die niet door de COTAN als voldoende is beoordeeld. Dat heeft tot gevolg dat de psycholoog in die gevallen testscores verzamelt waarvan de betekenis onvoldoende vaststaat. Het betekent in het beste geval dat de resultaten van ongeveer de helft van de testafnames weinigzeggend zijn, en in het slechtste geval dat de resultaten van ongeveer de helft van de testafnames misleidend zijn.’³⁵ De logica die uit deze conclusie spreekt valt enkel en alleen over-eind te houden als de COTAN een complete en geactualiseerde lijst van testbeoordelingen zou aanbieden. Maar daarvan is, herhalen we, in de verste verte geen sprake. De lijst met tests die de COTAN van een oordeel heeft voorzien is hoogst onvolledig en de testbeoordelingen die wel op deze lijst prijken zijn in veel gevallen gedateerd, dat wil zeggen ouder dan tien jaar. Het had de COTAN gesierd als ze het misverstand dat Schimmel en Van Koppen (2017) parten speelt – de COTAN als gouden standaard – zelf had weggenomen. Want nu worden de

29. B. Dandachi-FitzGerald e.a., ‘Verslag consensusbijeenkomst symptoomvaliditeitstests’, Utrecht 7 juni 2024.

30. Bijvoorbeeld W. Koops e.a., ‘Een duurzame toekomst voor de COTAN’, *De Psycholoog* 2025, 1, p. 20-25. De auteurs, (voormalig) leden van de COTAN, schrijven over hun commissie: ‘[...] deze NIP-commissie is van groot belang voor de toekomst van een wetenschappelijk testgebruik in Nederland. Ze functioneert als de belangrijkste verbindende schakel tussen praktiserende psychologen en onderzoekers’ (p. 20). En: ‘Hoezeer de COTAN inmiddels tot de maat der dingen was uitgegroeid, bleek wel bij de publicatie van deze nieuwe richtlijnen’ (p. 21).

31. ECLI:NL:RBAMS:2015:5279.

32. Bijvoorbeeld: ECLI:NL:RBROT:2022:9766. En zie voorts: ECLI:NL:GHDHA:2022:1748: ‘Appellant heeft verder nog gesteld – bij akte in hoger beroep – dat uit de testresultaten van de beoordeling [...] die hij bij COTAN heeft opgevraagd blijkt dat deze test een onvoldoende heeft gescoord op een testonderdeel en daarom niet voldoet aan het vereiste uit het Raamwerk dat de capaciteitentest dient te voldoen aan de COTAN-criteria.’

33. Volledigheidshalve merken we op dat de COTAN haar beperkingen ook wel ziet. De Commissie overweegt om een werkgroep van ter zake kundigen in te stellen, die moet bepalen welke beoordelingen dringend aan actualisering toe zijn. W. Koops e.a. 2025, p. 27: ‘Deze pro-actieve benadering moet zorgen voor relevantere en tijdigere beoordelingen.’

34. L. Schimmel & P. van Koppen, ‘Verdachten testen’, *De Psycholoog* 2017, 52(10), p. 20-23.

35. Schimmel & Van Koppen 2017, p. 40.

forensisch psychologen in het artikel van deze auteurs toch enigszins geportretteerd als stuntelige sukkels, die rechters van misinformatie voorzien. Het gevaar van zo'n soort sentiment is dat psychologen afzien van het gebruik van tests zolang deze tests niet met een gunstig oordeel op de COTAN-lijst prijken en zich in plaats daarvan baseren op hun klinische intuïtie. Dan raken we verder van huis.

In hun artikel drukken Schimmel en Van Koppen (2017) rechters op het hart om telkens deze vraag te stellen

aan psychologen wanneer die rapporteren in strafzaken en daarbij tests gebruiken: 'Wat is het oordeel van de COTAN over elke test die u gebruikt?'³⁶ Als het om validiteitstests gaat, kan het antwoord op die vraag luiden dat het oordeel van de COTAN eerder lijkt op dat van een oude Michelinids dan op dat van KEMA: onvolledig, gedateerd en daarmee irrelevant. Voor de onderbouwing van zo'n antwoord kan verwezen worden naar dit artikel.

36. Schimmel & Van Koppen 2017, p. 40.