

Forensisch psychologen spelen een belangrijke rol bij rechterlijke beslissingen over een verdachte. Lolotte Schimmel en Peter van Koppen doen verslag van een onderzoek naar de kwaliteit van de gebruikte tests in rapporten van forensisch psychologen. Beargumenteren psychologen waarom zij een bepaalde test gebruiken? Worden tests afgenomen onder toezicht of niet? 'In meer dan de helft van de gevallen wordt een test gebruikt die niet door de COTAN als voldoende is beoordeeld.'

## TESTGEBRUIK IN DE FORENSISCHE PSYCHOLOGIE

# VERDACHTEN TESTEN

### INLEIDING

Forensisch psychologen plegen hun rapport over een verdachte aan te heffen met 'Pro Justitia'. Die gewoonte – die overigens door niemand anders die bijdraagt aan het strafproces wordt gedeeld – is niet zonder betekenis. Het rapport van een forensisch psycholoog wordt niet geschreven met het oog op de belangen van de verdachte, maar als bouwsteen voor beslissingen in het strafproces.

De forensische rapporten hebben nog een ander kenmerk dat hen onderscheidt van typische psychologische rapportages: zij dragen bij aan een oordeel over de verdachte dat voor hem nogal verstrekkende gevolgen kan hebben. Op grond van zo'n pro Justitia-rapport kunnen tegen de verdachte allerlei dwangmaatregelen worden genomen. Andere psychologische rapporten zijn in de regel slechts een startpunt voor behandeling die gaande de rit nog kan worden bijgesteld.

Forensisch psychologen spelen derhalve een belangrijke en soms beslissende rol bij rechterlijke beslissingen over een verdachte (De Ruiter, 2004). Het gaat ook om forse aantallen. In 2015 werd in 698 strafzaken voor jeugdigen en in 2709 strafzaken voor volwassenen door forensisch psychologen rapport uitgebracht. Daarnaast leverde het Pieter Baan Centrum in 226 zaken een gezamenlijk rapport met onder andere de forensisch psychiater (NIPF, 2016). Het definitieve karakter van forensische rapporten en het belang dat zij dienen, stelt hoge eisen aan de kwaliteit van de rapporten. En die eisen moeten zeker worden gesteld aan de psychologische tests die bij verdachten worden afgenomen. Die zijn immers belangrijk voor de conclusies van de psycholoog (Ægisdóttir et al., 2006; Grove et al., Nelson, 2000).

Dit artikel is een verslag van een onderzoek naar de kwaliteit van de tests die worden gebruikt in rapporten van forensisch psychologen. Daarnaast is geprobeerd te achterhalen of (1) psychologen beargumenteren waarom zij een bepaalde test gebruiken, (2) tests worden afgenomen onder toezicht of niet en (3) of individuele psychologen een standaard testbatterij gebruiken of niet.

**DE FUNCTIE VAN HET TESTEN** In veel strafzaken vraagt de rechter-commissaris of de officier van Justitie aan een psycholoog om een rapportage op te stellen over de verdachte (Van Esch, 2012). Dat gebeurt in de regel naast een rapport van een forensisch psychiater. Psychiaters testen zelden, forensisch psychologen doen dat vrijwel altijd.

Het uitgangspunt is dat de psycholoog in het rapport een uitspraak doet over het recidiverisico en de mate van toerekeningsvatbaarheid van de verdachte ten tijde van het delict. Daartoe zou onder andere moeten worden onderzocht of de verdachte ten tijde van het misdrijf leed aan een gebrekkige ontwikkeling of psychiatrische stoornis en wat de mate van samenhang daarvan is met het ten laste gelegde misdrijf. Dat levert ten minste twee problemen op (Van Koppen, 2004). Ten eerste moet de psycholoog voor het onderzoek aannemen dat de verdachte het misdrijf pleegde. Ten tweede is het testen van iemands psychische toestand op een moment in het verleden, soms verre verleden, onmogelijk. Dat is een opgave die overigens niet wordt ondernomen. In plaats daarvan worden de tests gebruikt om de huidige toestand van de verdachte te onderzoeken.

De psycholoog verzamelt informatie met behulp van een instrumentarium waarvan de kenmerken, zoals betrouwbaarheid en validiteit, bekend zijn (Hoogerwerf et al., 2009). Het testpsychologisch onderzoek dient er om persoonskenmerken van de verdachte vast te stellen, pathologie te meten en om vast te stellen of er een gebrekkige ontwikkeling of ziekelijke stoornis is. De psycholoog gebruikt de tests ook om informatie uit andere bron, zoals de familie van de verdachte, te toetsen en om te onderzoeken of er discrepanties zijn met wat de verdachte zelf vertelt. Er zijn ook andere forensische relevante vragen, zoals het verband tussen de persoon en het ten laste gelegde en de zorg- en risicoprognose.

Het testgebruik werd in dit onderzoek beoordeeld op een tweetal relevante aspecten. Allereerst is onderzocht of tests worden gebruikt die van voldoende kwaliteit zijn. Daarnaast is onderzocht of tests worden gebruikt die geschikt zijn voor forensisch psychologische toepassing. Het eerste aspect werd onderzocht aan de hand van de testbeoordelingen van de Commissie Testaangelegenheden Nederland (COTAN) – een commissie van het Nederlands Instituut van Psychologen (NIP) die sinds 1959 de merites van psychologische tests onderzoekt (Drenth & Sijtsma, 2012). Het is de bedoeling dat de COTAN de meeste tests die in Nederland worden gebruikt een beoordeling geeft (Evers et al., 2010).

## Waarom heeft u deze tests bij deze verdachte gebruikt?

Het tweede aspect werd beoordeeld aan de hand van de zogenaamde *best practice* van het Nederlands Instituut voor Forensische Psychiatrie en Psychologie (NIFP; Hoogerwerf, et al., 2009). Het NIFP is een onderdeel van het ministerie van Veiligheid en Justitie. Vrijwel alle forensisch psychologische en forensisch psychiatrische rapporten worden gemaakt door medewerkers van het NIFP of door vrijgevestigden na bemiddeling door het NIFP. De rapporten werden vergeleken met de op het moment van rapporteren meest recente versie van de *best practice* van het NIFP uit 2009 (Hoogerwerf et al., 2009).

In dit artikel gaat het niet over de vraag of de interpretatie door de psycholoog juist is en of de conclusies in het individuele geval door de testscores worden gedragen. Wel zijn drie andere aspecten van de afname van de tests in het onderzoek betrokken. Dat gaat ten eerste over de vraag of

tests steeds onder toezicht worden afgenomen of dat tests worden meegegeven aan de verdachte. Ten tweede is onderzocht of psychologen steeds een vaste testbatterij gebruiken of steeds een wisselende. En ten derde is onderzocht of de psycholoog voor de gebruiker van het rapport de tests uitlegt en de keuze van tests beargumenteert.

De vraag naar toezicht bij de afname van een test is relevant voor het volgende. In de reguliere psychologische praktijk heeft degene die wordt onderzocht er meestal belang bij dat aan de psycholoog een zo juist mogelijk beeld van hem- of haarzelf wordt gegeven. In de forensische praktijk ligt dat anders (Rassin & Merckelbach, 2014). Verdachten willen soms zo gunstig mogelijk en in andere gevallen weer zo ongunstig mogelijk voor de dag komen. Het rapport van de forensisch psycholoog is immers van invloed op de straf en maatregel die gaat worden opgelegd. Als tests worden meegegeven, wordt zodoende ruim baan gegeven aan ongewenste invloeden op het testresultaat. Eén van de richtlijnen van het NIP is daarom dat er voortdurend toezicht of begeleiding is bij de afname van een test (Nederlands Instituut van Psychologen, 2010). Er zijn uitzonderingen, maar niet in de forensische praktijk. Volgens het Centraal Tuchtcollege Gezondheidszorg is er in het geval van een forensische situatie er in de regel een stroef onderzoekscontact dat het risico op problemen te groot maakt bij het meegeven van testmateriaal (Centraal Tuchtcollege voor de Gezondheidszorg, 2009).

Het is best practice rekening te houden met kenmerken van de verdachte bij de keuze voor een test (Hoogerwerf et al., 2009). Aan die aanbeveling zitten twee kanten. Ten eerste is het niet de bedoeling dat de forensisch psycholoog een soort standaardset aan tests gebruikt. Ten tweede moet er een zekere samenhang zijn met kenmerken van de persoon die wordt onderzocht en de gekozen tests. De psycholoog dient beperkingen van de verdachte mee te laten wegen bij de keuze voor een test, evenals een eventuele psychische stoornis, analfabetisme en een allochtone herkomst. Dat zou ertoe moeten leiden dat forensisch psychologen zich bedienen van wisselende testbatterijen. Het was niet mogelijk te onderzoeken of de gebruikte tests goed passen bij de verdachte. Als benadering daarvoor is gekozen om te onderzoeken of de forensisch psychologen steeds een vaste testbatterij gebruiken of niet.

### TESTBEOORDELINGEN DOOR DE COTAN

De COTAN beoordeelt psychologische tests aan de hand van zeven criteria (Evers et al., 2010):

1. *De uitgangspunten van de testconstructie*, namelijk of het gebruiksdoel, de theoretische achtergrond en de operationalisatie ervan helder zijn beschreven.
2. *De kwaliteit van het testmateriaal*, namelijk of de testopgaven, de scoring en de instructie zijn gestandaardiseerd en of er genoeg aanwijzingen voor de verdachte gegeven worden.
3. *De kwaliteit van de handleiding*, namelijk of de informatie in de handleiding volledigheid biedt voor de gebruiker.
4. *De normen van de test*. Als er geen normscores zijn, is het oordeel in principe onvoldoende, met enige uitzonderingen.
5. *De betrouwbaarheid*, waarbij men een oordeel geeft over de betrouwbaarheidsanalyse en het uitgevoerde onderzoek naar de betrouwbaarheid.
6. *De begripsvaliditeit*, namelijk of de test de eigenschap meet waarvoor de test bedoeld is.
7. *De criteriumvaliditeit*, namelijk in hoeverre de testscore een goede voorspeller is van wat de test beoogt te meten.

Op elk criterium kan de beoordeling 'onvoldoende', 'voldoende' of 'goed' zijn (Evers et al., 2010). Een onvoldoende betekent ofwel dat het criterium wel is onderzocht maar dat de kwaliteit onvoldoende is, ofwel dat het criterium niet is onderzocht. Een onvoldoende op een of meer criteria van een test hoeft niet te betekenen dat een test onbruikbaar is. Een onvoldoende heeft voor de gebruiker vooral een functie als waarschuwingssignaal. De gebruiker moet in een dergelijk geval goed kunnen beargumenteren waarom hij toch een bepaalde test inzet.

Voor het onderzoek was de COTAN-beoordeling leidend. Daarbij is echter slechts gelet op de beoordeling van normen, betrouwbaarheid en begrips- en criteriumvaliditeit, indien bekend, omdat die het belangrijkste zijn voor de voorspellende waarde van de tests (COTAN-documentatie; Hoogerwerf et al., 2009; Kamphuis & Geurts, 2006). Als op drie van die vier criteria minimaal een voldoende is gescoord, is de kwaliteit als voldoende beschouwd. In de andere gevallen was het oordeel onvoldoende. Met de uitdrukking dat een test voldoende of onvoldoende is, wordt verder steeds bedoeld op die samenvatting van het COTAN-oordeel.

#### BEST PRACTICE VAN HET NIFP

Het NIFP geeft een aanbeveling wat de best practice zou zijn in forensisch psychologische rapportages (Hoogerwerf et al., 2009). Aan het testonderzoek wordt in de best practice veel aandacht besteed. Zo geeft het NIFP aanbevelingen hoe het

testpsychologisch onderzoek zou moeten worden uitgevoerd en wat daarover zou moeten worden gerapporteerd. Omdat de opdrachtgevers in het forensische veld – advocaten, officieren van justitie en rechters – niet bekend zijn met psychologische tests, is het best practice om een omschrijving van de gebruikte tests te geven. Daarom moeten ook de keuze en het gebruik van tests wordt verantwoord tegenover de opdrachtgevers. Het is ook best practice om eventuele beperkingen van de verdachte mee te laten wegen bij de keuze voor een test, evenals een eventuele psychische stoornis, analfabetisme en een allochtone herkomst (Hoogerwerf et al., 2009).

Het is best practice dat er eerst onderzoek wordt gedaan naar de cognitieve en functionele capaciteiten van de verdachte. Vervolgens kan er onderzoek worden verricht naar relevante persoonskenmerken en diagnostische classificaties volgens de *Diagnostic and Statistical Manual of Mental Disorders* (DSM; American Psychiatric Association, 2013).

In de best practice geeft het NIFP voorts aanbevelingen voor tests die gebruikt kunnen worden. Er worden niet expliciet bepaalde tests afgeraden. Waar in de rest van het artikel gesproken wordt over aanbevolen en niet-aanbevolen, wordt steeds bedoeld wat daarover in de best practice wordt gezegd. De best practice geeft richtlijnen voor de tests die in elk domein moeten worden gebruikt.

## Psychiaters testen zelden, forensisch psychologen doen dat vrijwel altijd

#### METHODE

**STEELPROEF** Er werd een steekproef getrokken uit forensisch psychologen die zijn geregistreerd door het Nederlands Register Gerechtelijk Deskundigen (NRGD). Forensisch psychologen die willen worden geregistreerd, moeten naast een curriculum vitae en een aantal andere documenten een aantal schriftelijke rapporten inleveren. De rapporten moeten overigens vooraf worden geanonimiseerd. Een belangrijk criterium voor het toelaten tot het register is de kwaliteit van de ingeleverde rapporten zodat kan worden aangenomen dat de forensisch psychologen vooral hun beste rapporten zullen inleveren.

Forensisch psychologen kunnen worden geregistreerd als rapporteur Jeugd en rapporteur Volwassenen. De steekproef werd beperkt tot forensisch psychologen die zijn geregistreerd voor volwassenen. Op het moment van het onderzoek waren in die categorie 192 personen geregistreerd. Wij hadden geen toegang tot de cv's van de psychologen wiens rapporten in de steekproef zitten. Wel was bekend dat zij allen geregistreerd zijn bij het NRGD en derhalve door de selectieprocedure zijn gekomen.

De aselechte steekproef omvatte 20% van die populatie, dat wil zeggen 38 psychologen. Van die psychologen zijn alle ingeleverde rapporten die ook echt over volwassenen gingen in het onderzoek betrokken. Dat waren 175 rapporten. In 10 rapporten (6%) heeft er geen testonderzoek plaatsgevonden omdat de verdachte weigerde mee te werken, zodat het onderzoek is uitgevoerd op 165 rapporten. Vrijwel alle rapporten zijn geschreven in de jaren 2009 tot en met 2015. Drie werden geschreven in 2008 en één in 2016.

Alle rapporten werden door de eerste auteur gecodeerd. Er is geen tweede codeur ingezet, omdat het codeerwerk weinig risico op fouten had. Het ging immers steeds om de eenvoudige codering of een test is ingezet of niet.

## RESULTATEN

**COTAN EN BEST PRACTICE** Er is in de 165 rapporten in totaal 1074 keer een test afgenomen, gemiddeld 6,5 per onderzochte verdachte. De forensisch psychologen gebruikten liefst 101 verschillende tests – zie tabel 1 voor de meest gebruikte tests. Tabel 2 vat het oordeel van de best practice en de COTAN samen. In 168 van de afnames (16%) was het een projectieve test die niet voldoet aan basale eisen van validiteit (Ben-Shakar et al., 1998; Grove et al., 2002).

Psychologen volgen niet altijd de door het NIPF aanbevolen best practice. Van de 101 afgenomen tests komt 57% niet voor in de best practice. De aanbevolen tests worden wel vaker gebruikt, zodat van de testafnames 85% conform best practice is.

De *proof of the pudding* is het oordeel van de COTAN. Minder dan de helft van testafnames, 47%, gebeurde met een door de COTAN getoetst en voldoende bevonden test. Dat percentage varieerde van rapporteur tot rapporteur, en wel tussen 18,5% en 81,3%. Van de 101 verschillende tests die worden gebruikt, is 28% volgens COTAN-normen van voldoende kwaliteit.

De getallen voor best practice-tests en COTAN-goedgekeurde-tests lopen uiteen omdat in de NIPF best practice nogal eens tests worden aanbevolen die van onvoldoende kwaliteit

zijn (zie tabel 2). Van de 43 tests die in de best practice worden aanbevolen, zijn er slechts 19 (44%) door de COTAN als van voldoende kwaliteit beoordeeld, minder dan de helft derhalve.

**UITLEG EN BEGELEIDING TESTAFNAME** Het is best practice dat psychologen een beschrijving van de test geven en hun keuze beargumenteren in hun rapport. Dat gebeurt weinig. In 5,2% van de rapporten (n=56) beargumenteerde de psycholoog in het rapport de keuze van de test en in minder dan de helft (46,2%; n=496) werd een beschrijving van de test gegeven.

Er dient bij afname van testmateriaal in een forensische situatie begeleiding en toezicht te zijn (College van Beroep van het Nederlands Instituut voor Psychologen, 1999). Uit de forensische rapporten blijkt in de meeste gevallen niet of aan die eis is voldaan. Slechts zelden wordt precies beschreven hoe de testafname is gebeurd en wie daarvoor verantwoordelijk was. In een deel van de rapporten werd zonder meer opgeschreven dat de tests aan de verdachte waren meegegeven om die zelfstandig, zonder begeleiding en toezicht, in te vullen. Dat gold voor 4,7% (n=50) van de tests. Voor nog eens 8,8% (n=95) van de tests suggereerden de formuleringen in de rapporten dat zij aan de verdachte zijn meegegeven. In de meeste gevallen werd er niet over begeleiding of toezicht gerapporteerd.

**AANPASSEN AAN VERDACHTE** Het onderzoek bood te weinig inzicht in de verdachten om te bepalen of de gebruikte tests steeds goed gekozen waren. Als benadering is gekozen voor aanpassing van de tests aan de verdachte en de variatie in de tests die een psycholoog gebruikt, dat wil zeggen het percentage tests dat bij rapporten van elke psycholoog steeds hetzelfde was. De mate van overeenkomst in testgebruik tussen verschillende rapporten van dezelfde psycholoog zegt meer als er van een psycholoog negen rapporten beschikbaar waren dan als er slechts twee rapporten waren. Naar verwachting neemt de mate van overeenkomst af bij een toename van het aantal afgenomen testbatterijen. Er zou derhalve een negatief verband moeten zijn tussen het aantal testbatterijen en de mate van overeenkomst tussen testbatterijen.

Voor dat negatieve verband zijn echter geen aanwijzingen gevonden. Het tegenovergestelde is het geval – zie tabel 3. Door 25 van de 38 psychologen werd bij elk onderzoek steeds meer dan 50% dezelfde tests gebruikt. Zo hebben achttien psychologen vier rapporten geschreven waarvan de gemid-

TABEL 1: OVERZICHT MEEST GEBRUIKTE TESTS\* (N&gt;10) IN 165 RAPPORTEN (N=1074).

	N	% VAN AANTAL RAPPORTEN	NIFP BEST PRACTICE	COTAN-BEOORDELING
<b>INTELLIGENTIE TESTS<sup>1</sup></b>				
Wechsler Adult Intelligence Scale-III (WAIS-III)	76	46,1	Verouderd <sup>2</sup>	onvoldoende
WAIS-IV	30	18,2	Aanbevolen	voldoende
Verkorte GIT-2 <sup>3</sup>	23	13,9	Aanbevolen	–**
Kaufman – Intelligentietest	13	7,9	Aanbevolen	voldoende
Raven Standard Progressive Matrices	13	7,9	Bruikbaar voor anderstaligen	voldoende
<b>OVERIGE (PERSOONLIJKHEIDS)TESTS</b>				
Utrechtse Copinglijst	97	58,8	Aanbevolen	voldoende
Minnesota Multiphasic Personality Inventory (tweede versie)	96	58,2	Aanbevolen om standaard deel uit te maken van testbatterij	onvoldoende
Nederlandse Persoonlijkheidsvragenlijst	83	50,3	Aanbevolen als aanvulling op NVM of MMPI-2	voldoende
Nederlandse Verkorte MMPI	83	50,3	Aanbevolen om standaard deel uit te maken van testbatterij	voldoende
Symptom CheckList (SCL-90)	54	32,7	Aanbevolen als aanvulling op NVM of MMPI-2	voldoende
Buss-Durkee Hostility Inventory-Dutch	47	28,5	Aanbevolen	voldoende
Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)	13	7,9	Aanbevolen, maar minder geschikt bij forensisch gebruik	–
Structured Inventory of Malingered Symptomatology	13	7,9	Aanbevolen	onvoldoende
Delis-Kaplan Executive Function System Trail Making Test	11	6,7	Aanbevolen	onvoldoende
Negativisme, Ernstige psychopathologie en somatisteringstest	11	6,7	–	voldoende
<b>PROJECTIEVE TESTS</b>				
Zinnen Aanvul Test	79	47,9	Projectieve test als aanvulling op gestandaardiseerde tests	onvoldoende
House, Tree, Person	36	21,8	Projectieve test als aanvulling op gestandaardiseerde tests	–
Thematic Apperception Test	35	21,2	Projectieve test als aanvulling op gestandaardiseerde tests	–
<b>TOTAAL</b>	<b>1074</b>	<b>100% (=165)</b>		

1: Best practice: in elk rapport moet onderzoek worden gedaan naar intelligentie.

2: Er is vanaf 2012 een nieuwe versie van de WAIS. De best practice stamt van voor die tijd. De rapporten stammen uit 2008 tot 2016 en er is te zien dat het gebruik van de WAIS-III op een bepaald moment wordt vervangen door de WAIS-IV.

3: Een verkorte GIT-2 (6 subtests) wordt aanbevolen. Een sterk verkorte

(3 subtests) of verouderde versie wordt niet aanbevolen. De COTAN beoordeelt de sterk verkorte versie als onvoldoende; een COTAN beoordeling van een verkorte GIT-2 is er niet.

\* Een tabel met alle gebruikte tests is opvraagbaar bij de auteurs.

\*\* – betekent dat de test niet is genoemd in de best practice of dat hij niet door de COTAN is beoordeeld. Zie voor de literatuurverwijzingen bij de tests de documentatie van de COTAN.

delde overeenkomst tussen de testbatterijen 55% was. Dat geeft een indicatie voor het feit dat psychologen een voorkeur hebben voor het gebruik van dezelfde tests, maar voor een deel het testgebruik ook afstemmen op de verdachte. De gebruikte methode geeft echter te weinig inzicht in de vraag of testbatterijen worden aangepast aan de onderzochte verdachte. De getallen vormen niet meer dan een aanleiding om dit probleem nader te onderzoeken.

## DISCUSSIE

Het doel van het onderzoek was te inventariseren wat de kwaliteit is van het gebruik van psychologische tests in forensisch psychologische rapportages. Er is onderzocht welke tests worden gebruikt en wat volgens de COTAN de kwaliteit van die tests is. Ook is onderzocht of de psychologen volgens de best practice-aanbevelingen testonderzoek doen en of zij hun psychologisch testonderzoek afstemmen op de verdachte.

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. Dat vormt een slechte basis voor rapportages over verdachten aan de rechter.

In het onderzoek bleek overigens dat in de best practice van het NIPP nogal eens tests worden aanbevolen die niet door de COTAN als voldoende zijn beoordeeld. Het is onduidelijk welk ander criterium is gebruikt voor de *best practice* aanbevolen tests dan de oordelen van de COTAN. Bij een onvoldoende op een of meer van de criteria van de COTAN zou de gebruiker behoren te beargumenteren waarom hij toch voor die test kiest (Evers et al., 2010). Voor de best practice aanbevolen tests ontbreekt een dergelijke argumentatie bij de afzonderlijke tests (Hoogerwerf et al., 2009). Men mag verwachten dat in een dergelijk leidend stuk wordt beargumenteerd waarom het gebruik van elke genoemde test best practice is, juist als zij onvoldoende scoren op een of meer criteria van de COTAN en dan in het bijzonder op de criteria validiteit, normen en betrouwbaarheid. Dat zijn namelijk belangrijke psychometrische eigenschappen die wat zeggen over de waarde van de testuitslagen bij de verdachten (Drenth & Sijsma, 2012). Verwacht zou mogen worden dat de best practice de beste test in zijn soort aanbeveelt. Dat is nu niet duidelijk.

Het onderzoek laat eveneens zien dat forensisch psychologen zelden een duidelijke reden geven voor hun keuze van tests en in het merendeel van de gevallen geen beschrijving van de gekozen test aan de gebruikers van het rapport aanleveren. Dat is niet in overeenstemming met de best

TABEL 2: SAMENVATTING OORDEEL BEST PRACTICE EN COTAN (N=1074).

OORDEEL BEST PRACTICE	OORDEEL COTAN	AANTAL	PERCENTAGE
Niet aanbevolen	Niet onderzocht	93	8,7
Niet aanbevolen	Onvoldoende	40	3,7
Niet aanbevolen	Voldoende	33	3,1
Aanbevolen	Niet onderzocht	125	11,6
Aanbevolen	Onvoldoende	309	28,8
Aanbevolen	Voldoende	474	44,1
Projectieve test		168	15,6
Aanbevolen		908	84,5
Niet aanbevolen		166	15,5
	Voldoende	507	47,2
	Onvoldoende	349	32,5
	Niet onderzocht	218	20,3
TOTAAL		1074	100



practice-aanbevelingen (Hoogerwerf et al., 2009). De gebruikers zijn in de forensische praktijk immers vrijwel altijd leken.

Hoewel de psychologen in het merendeel van de gevallen zelf het testonderzoek afnemen, zoals behoort in een forensische situatie (Centraal Tuchtcollege voor de Gezondheidszorg, 2009), worden de tests nogal eens meegegeven aan de verdachte om thuis of in de cel al dan niet alleen te maken. De betekenis van de resultaten van dergelijke testafnames zijn nihil, omdat dan niet meer duidelijk is of de test bij de verdachte is afgenomen of bij iemand anders.

Tot slot probeerden wij met de variatie in gekozen testbatterijen te benaderen of psychologen hun tests aanpassen aan de verdachte. Daarin zijn wij onvoldoende geslaagd. Er is wel zichtbaar dat de variatie, tegen de verwachting in, geringer is als de psycholoog veel rapporten inleverde, maar de betekenis daarvan is niet duidelijk. Er kan namelijk niet uitgesloten worden dat sommige veel rapporterende psychologen juist gespecialiseerd zijn in een bepaald type verdachte en daardoor vaker dezelfde testbatterij gebruiken. Dit punt wordt daarom verder buiten de beschouwing gelaten.

Het onderzoek laat zien dat een belangrijke bouwsteen van rapporten door forensisch psychologen in Nederland te wensen overlaat. Dat brengt het risico met zich mee dat in rapporten gebaseerd op tests van onvoldoende kwaliteit, en geringe toelichting, de kwaliteit van de conclusies evenredig geringer is. Als op dergelijke conclusies belangrijke beslissingen worden genomen over de verdachte, geeft dat geen garantie voor de juistheid van de conclusies en de beslissin-

gen van de rechter. Het ware daarom voor rechters verstandig om aan forensisch psychologen, naast de gebruikelijke vragen, steeds het volgende voor te leggen:

1. Wat is het oordeel van de COTAN over elke test die u heeft gebruikt?
2. Past elke test die u gebruikte en de manier waarop die is gebruikt in de best practice-aanbevelingen van het NIPP?
3. Waarom heeft u deze tests bij deze verdachte gebruikt?
4. Zijn de tests onder toezicht afgenomen?

Pas bij een bevredigend antwoord op die vragen zouden de conclusies van de forensisch psycholoog over de verdachte mogen worden aanvaard.

#### OVER DE AUTEURS

Lolotte M. C. Schimmel is student aan de Faculteit der Rechtsgeleerdheid & de Faculteit der Gedrags- en Bewegingswetenschappen, Vrije Universiteit Amsterdam. E-mail: [lmcschimmel@gmail.com](mailto:lmcschimmel@gmail.com). Prof. dr. Peter J. van Koppen is hoogleraar Rechtspsychologie aan de Faculteiten der Rechtsgeleerdheid, Vrije Universiteit Amsterdam en Universiteit Maastricht. Hij is ook lid van het College van het Nederlands Register Gerechtelijk Deskundigen. E-mail: [p.j.van.koppen@vu.nl](mailto:p.j.van.koppen@vu.nl).

De auteurs danken M. Vere van Koppen en Jan de Keijser en twee reviewers voor hun commentaar op eerdere versies van dit artikel. Wij danken ook het bureau van het Nederlands Register voor Gerechtelijke Deskundigen (NRGD) voor hun hulp en gastvrijheid bij het verzamelen van de gegevens.

TABEL 3: GEMIDDELD PERCENTAGE GELIJKE TESTS.

AANTAL RAPPORTEN PER PSYCHOLOOG	AANTAL PSYCHOLOGEN	GEMIDDELD PERCENTAGE GELIJKE TESTS
1	1	
2	2	39
3	7	55
4	18	55
5	2	54
6	3	53
7	2	64
8	2	57
9	1	84
	38	



## Een belangrijke bouwsteen van rapporten door forensisch psychologen laat in Nederland te wensen over

### Summary

TESTING SUSPECTS: USE OF TESTS BY FORENSIC PSYCHOLOGIST

L.M.C. SCHIMMEL & P.J. VAN KOPPEN

The quality of tests used in experts reports by forensic psychologist in The Netherlands in criminal cases was studied. Tests were assessed using the so-called *best practice* by the Netherlands Institute for Forensic Psychiatry and Psychology (NIFP) and the Committee on Test Matters (COTAN) of the

Dutch Institute for Psychologists (NIP). From the forensic psychologists registered by the Netherlands Registry of Court Experts (NRGD) 20% (N=38) was sampled and all their reports (165) submitted for their registration were studied. In total 1074 times a test was used in these 165 reports. Most of them were best practice recommended, but only 47% were judged to be of any quality by the COTAN. In a small minority, the psychologists

explain their use of tests to the legal users of their reports. Rather frequently, suspects are given the opportunity to make tests without supervision, at home or in their cell, making the results dubious. It is concluded that the use of tests in a large number of forensic psychological expert reports does not meet standards. Courts should always critically interrogate these experts about the use of psychological tests.

### Literatuur

- Ægisdóttir, S., White, M.J., Spengler, P.M., Maugherman, A.S., Anderson, L.A. et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *Counseling Psychologist*, 34, 341-382.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders: Fifth edition (DSM-V)*. Arlington, VA: American Psychiatric Publishing.
- Ben-Shakhar, G., Bar-Hillel, M., Bilu, Y. & Shefler, G. (1998). Seek and you shall find: Test results are what you hypothesize they are. *Journal of Behavioral Decision Making*, 11, 235-249.
- Centraal Tuchtcollege voor de Gezondheidszorg (2009). *Beslissing in de zaak onder nummer 2007/300*.
- College van Beroep van het Nederlands Instituut voor Psychologen (1999). College van Beroep van het Nederlands Instituut voor Psychologen, 20 juni 1997, (Psycholoog in Ballpenmoordzaak). *Tijdschrift voor Gezondheidsrecht*(5), 317-321.
- Cotan-documentatie. <http://www.cotandocumentatie.nl/> (geraadpleegd op 26 januari 2017).
- Drenth, P.J.D. & Sijtsma, K. (2012). *Testtheorie: Inleiding in de theorie van de psychologische test en zijn toepassingen* (4e ed.). Houten: Van Loghum Slaterus.
- Esch, C. van (2012). *Gedragsdeskundigen in strafzaken*. Assen: Van Gorcum (Dissertatie Universiteit Leiden).
- Evers, A., Lucassen, W., Meijer, R. & Sijtsma, K. (2010). *Cotan beoordelings-systeem voor de kwaliteit van tests*. Utrecht: Nederlands Instituut van Psychologen.
- Grove, W.M., Barden, R.C., Garb, H.N. & Lilienfeld, S.O. (2002). Failure of Rorschach-Comprehensive-System-based testimony to be admissible under the Daubert-Joiner-Kumho standard. *Psychology, Public Policy, and Law*, 8, 216-234.
- Grove, W.M., Zald, D., Lebow, B., Snitz, B. & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12, 19-30.
- Hoogerwerf, R., Kordelaar, W. van, Pauw, J., Verheugt, T. & Woudenberg, I. van (2009). *Best practice ambulante: Forensisch onderzoek en rapportage in het strafrecht voor volwassenen*. Utrecht: Nederlands Instituut voor Forensische Psychiatrie en Psychologie.
- Kamphuis, J. & Geurts, H. (2006). Gestandaardiseerde psychodiagnostische methoden. In W. Vandereycken, C. Hoogduin & P. Emmelkamp (Eds.), *Handboek psychopathologie. Deel 2 klinische praktijk*. Houten: Bohn Stafleu van Loghum.
- Koppen, P.J. van (2004). Weg van de toerekeningsvatbaarheid: Over rapportages over de verdachte. *Trema*, 27, 221-228.
- Nederlands Instituut van Psychologen (2010). *Algemene standaard testgebruik*. Utrecht: NIP.
- NIFP (2016). *Nederlands Instituut voor Forensische Psychiatrie en Psychologie: Jaarbericht 2015*. s.l.: Auteur.
- Rassin, E. & Merckelbach, H. (2014). Forensisch-psychiatrische diagnostiek: Hoog tijd om er een echt vak van te maken. *Nederlands Juristenblad*, 2980-2986.
- Ruiter, C. de (2004). Forensisch gedragsonderzoek in strafzaken. *Justitiële Verkenningen*, 30(1), 50-60.
- Sijtsma, K. & Emons, W.H.M. (2007). Korte tests: Kostbare tijdwinst en onbetrouwbare beslissingen. *De Psycholoog*, 42, 406-411.

