

Bedankt voor het downloaden van dit artikel. De artikelen uit de (online)tijdschriften van Uitgeverij Boom zijn auteursrechtelijk beschermd. U kunt er natuurlijk uit citeren (voorzien van een bronvermelding) maar voor reproductie in welke vorm dan ook moet toestemming aan de uitgever worden gevraagd.

# Boom

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch door fotokopieën, opnamen of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van kopieën uit deze uitgave is toegestaan op grond van artikelen 16h t/m 16m Auteurswet 1912 jo. Besluit van 27 november 2002, Stb 575, dient men de daarvoor wettelijk verschuldigde vergoeding te voldoen aan de Stichting Reprorecht te Hoofddorp (postbus 3060, 2130 KB, [www.reprorecht.nl](http://www.reprorecht.nl)) of contact op te nemen met de uitgever voor het treffen van een rechtstreekse regeling in de zin van art. 16l, vijfde lid, Auteurswet 1912.

Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16, Auteurswet 1912) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten, postbus 3060, 2130 KB Hoofddorp, [www.cedar.nl/pro](http://www.cedar.nl/pro)).

*No part of this book may be reproduced in any way whatsoever without the written permission of the publisher.*

[info@boomamsterdam.nl](mailto:info@boomamsterdam.nl)  
[www.boomuitgeversamsterdam.nl](http://www.boomuitgeversamsterdam.nl)

# De hiërarchische Cognitieve Screening Test: Constructvaliditeit en test-hertestbetrouwbaarheid

► Han Diesfeldt

■ **Samenvatting** — De Cognitieve Screening Test (CST) is een instrument voor initiële screening op een neurocognitieve stoornis. Negen van de twintig items vormen een in moeilijkheidsgraad oplopende, hiërarchische schaal. Deze korte hiërarchische variant en de complete versie van de CST zijn in twee sessies voorgelegd aan 65 ouderen die poliklinisch neuropsychologisch werden onderzocht wegens het vermoeden van een neurocognitieve stoornis. De eerder vastgestelde moeilijkheidshiërarchie van negen items werd bevestigd. De verschillende varianten van de CST correleren hoog met de resultaten van een andere veelgebruikte geheugentest, de Visuele Associatie Test. Uit test-hertestonderzoek met de korte variant (CST9) zijn referentiegegevens afgeleid om in herhaald onderzoek bij individuele patiënten betekenisvolle scoreveranderingen te kunnen vaststellen.

## Inleiding

Met de CST kunnen onderzoekers nagaan in welke mate een onderzochte persoon eenvoudige feitenkennis vanuit het geheugen kan ophalen (De Graaf & Deelman, 1991). Uit de antwoorden kan een hypothese over aanwezigheid van een neurocognitieve stoornis worden afgeleid (Van Toutert e.a., 2016). De CST is opgenomen in diverse richtlijnen en overzichten voor het gebruik van cognitieve meetinstrumenten (NIP, 2010; Pastink e.a., 2019; Wattel & Achterberg, 2010).

Eerder is vastgesteld dat negen items uit de CST een betrouwbare weergave vormen van de complete versie (Diesfeldt, 2012). De negen items zijn geselecteerd met Mokken-schaalanalyse, een techniek uit de non-parametrische item-responstheorie (Sijtsma & Van der Ark, 2017). De items van de CST9 vormen een homogene, hiërarchische schaal: zij zijn te ordenen naar moeilijkheidsgraad. In een hiërarchische schaal worden de items een voor een, in opklimmende graad van moeilijkheid aan een onderzochte persoon voorgelegd. In theorie kan de testafname

stoppen na het eerste item in de hiërarchie dat een onderzochte niet goed kan beantwoorden. De score is dan gelijk aan het rangnummer van het laatste item dat goed werd beantwoord ofwel de somscore van de goed beantwoorde voorafgaande items (Stochl e.a., 2012; Streiner e.a., 2015). Deze werkwijze heeft als voordeel dat onderzochten niet verder worden belast met items waarop zij hoogstwaarschijnlijk het antwoord schuldig blijven. De wetenschappelijke literatuur telt tal van voorbeelden van hiërarchisch geordende schalen die met behulp van Mokken-schaalanalyse zijn geconstrueerd (McGrory e.a., 2015; Watson e.a., 2012).

De CST9 heeft een scorebereik van 0 tot 9. In een onderzoek bij 102 ouderen met een normale cognitie behaalde geen enkele deelnemer een score lager dan 7, en 95,1% een hogere score (Van Toutert e.a., 2016). Scores > 7 kunnen daarom gelden als 'normaal' (Guilmette e.a., 2020; Hendriks e.a., 2020).

Onderzoek naar de constructvaliditeit van de CST9 bij 642 deelnemers aan psychogeriatrische dagbehandeling liet een Pearson-correlatie van 0,60 zien met een combinatie van drie subtests uit de Amsterdamse Dementie-Screeningstest (ADS3): Visueel Geheugen, Oriëntatie en Fluency (Diesfeldt, 2012). De CST9 was in dit onderzoek afgeleid uit de complete variant, CST20. De validiteit van een hiërarchisch geconstrueerde schaal moet zich bewijzen in nieuw onderzoek (Watson e.a., 2008). Systematische citatieanalyse van studies over hiërarchische schalen leverde echter geen voorbeelden op van dergelijk valideringsonderzoek.

Voor een hiërarchisch geconstrueerde schaal kan op twee manieren een score worden bepaald: over alle items, of wanneer zou zijn gestopt na het eerste onjuist beantwoorde item (zie Box 1). Stoppen na het eerste onjuist beantwoorde item kan een lagere score opleveren dan afname van alle items. Een hiërarchische schaal berust namelijk op een waarschijnlijkheidsmodel, berekend voor een grote groep deelnemers. De opklimende graad van moeilijkheid hoeft echter niet voor iedere individuele deelnemer te gelden. Het kan dus voorkomen dat een deelnemer een relatief gemakkelijk item niet kan beantwoorden, maar een moeilijker item met een hoger itemnummer wel (Stochl e.a., 2012).

### Box 1 Inhoud van de Cognitieve Screening Test

De Cognitieve Screening Test (CST20) telt twintig items die een beroep doen op episodische en semantische geheugenkennis. Episodische geheugenkennis is sterk aan een temporeel veranderlijke context gebonden en wordt getest met vragen naar de precieze datum, de dag

van de week, de eigen leeftijd, hoe laat het is en het jaargetijde. Semantische geheugenkennis wordt getest met vragen naar woonadres, geboortedatum, regerend vorst, minister-president, hun ambtsvoorgangers en de jaartallen van de twee wereldoorlogen (Diesfeldt, 2003; 2006). Negen items met een opklimmende graad van moeilijkheid vormen een korte, hiërarchische variant: CST9. Bij afname van een hiërarchisch opgebouwde schaal kan in theorie gestopt worden bij het eerste item waarop een onderzochte geen antwoord weet. De score is dan gelijk aan de somscore van de goed beantwoorde voorafgaande items. Stoppen na het eerste onjuiste antwoord, hier aangeduid als CST9a, wordt in dit artikel nader onderzocht en vergeleken met afname van de complete CST20 en alle negen items van de CST9.

De negen CST-items werden twee keer afgenomen: 1) als onderdeel van de complete CST20, en 2) zelfstandig, als de korte, hiërarchische vorm (CST9). In dit nieuwe onderzoek met de complete en de hiërarchische CST-variant kunnen dan ook drie vragen worden beantwoord: 1) Levert stoppen na het eerste onjuiste antwoord in de hiërarchische variant (CST9a) dezelfde score op als wanneer alle negen vragen worden voorgelegd (CST9)? 2) Wat is de constructvaliditeit en 3) de test-hertestbetrouwbaarheid van de CST9?

## Methode

### *Patiënten*

In dit tijdschrift vroegen we gebruikers van de CST20 om deel te nemen aan nieuw onderzoek naar de validiteit van de korte, hiërarchische variant van de CST, de CST9 (Van Toutert e.a., 2016). Zeven psychologen toonden belangstelling, vier van hen leverden gegevens aan. Zij gebruikten voor het reguliere neuropsychologisch onderzoek minstens twee sessies, zodat bij aanvang van de ene sessie de complete CST20 kon worden afgenomen en bij aanvang van de andere sessie de korte, hiërarchische variant (CST9). Aan de onderzoekers werd gevraagd om de twee volgordes 'complete variant-korte variant' en 'korte variant-complete variant' af te wisselen. Verder leverden de onderzoekers gegevens aan uit de voor hen gebruikelijke testbatterij.

### *Werkwijze*

De deelnemende psychologen verzamelden van 2017 tot 2020 gegevens van 65 patiënten (34 vrouwen, 31 mannen). De patiënten werden onder-

zocht vanwege vermoeden van een neurocognitieve stoornis. Naast de CST kregen 55 patiënten de Visuele Associatie Test (VAT, vorm A). Andere tests, waaronder de Achtwoordentest (dertig), MoCA Montreal Cognitive Assessment (dertien) en ADS3 (vier) waren bij veel minder patiënten afgenomen (tussen haakjes de aantallen). Voor validering van scores op de diverse CST-varianten is daarom de VAT gebruikt.

## Resultaten

Gegevens over leeftijd en opleiding van de deelnemers staan in Tabel 1, evenals het interval tussen de twee testmomenten en het gemiddelde op de diverse CST-varianten. Afname van alle negen items (CST9) leverde op groepsniveau een hoger gemiddelde op dan wanneer zou zijn gestopt na het eerste onjuiste antwoord (CST9a).

TABEL 1 Deelnemerskenmerken (N = 65) en testresultaten

Variabele	Bereik	Mediaan	Gemiddelde (SD)
Leeftijd	54-94	80	79,6 (10,2)
Opleiding <sup>a</sup>	1-7	4	4,1 (1,6)
Dagen tussen twee testmomenten	1-28	7	8,6 (5,0)
CST20 (standaardversie)	5,5-20	15,5	14,4 (4,2)
CST9 (hiërarchische variant)	0-9	7	6,7 (2,5)
CST9a (stop na eerste onjuiste antwoord)	0-9	6	5,7 (2,9)

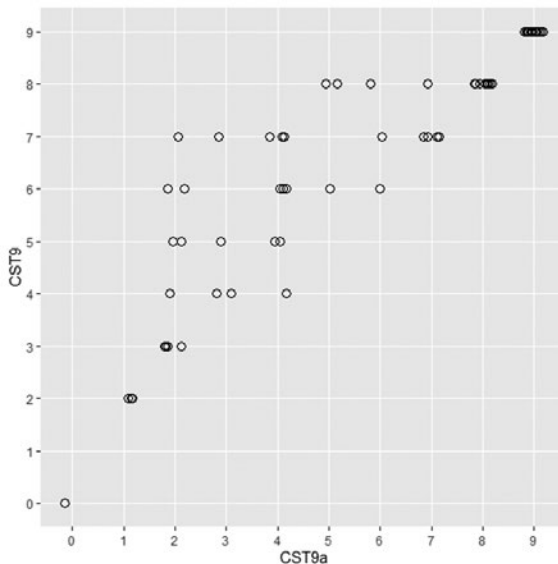
Afkortingen: CST = Cognitieve Screening Test; SD = standaarddeviatie; a Volgens de classificatie van Verhage, 1964

De volgorde waarin de CST20 en de CST9 werden afgenomen was niet gelijk verdeeld. Veel vaker (48 keer) werd de CST20 in de eerste sessie afgenomen dan de CST9 (zeventien keer). Om na te gaan in hoeverre afnamevolgorde invloed had op het verband tussen de CST20- en de CST9-score werd een regressieanalyse uitgevoerd met de CST20-score en testvolgorde als onafhankelijke variabelen. Testvolgorde had geen invloed op de voorspelde CST9-score, noch op de regressiecoëfficiënten. Vergeleken met een regressiemodel zonder de variabele 'testvolgorde' had het model met 'testvolgorde' geen toegevoegde verklarende waarde ( $F_{2,61} = 2,13$ ;  $p = 0,13$ ).

### Bevestiging van de rangordening van items naar moeilijkheidsgraad

Tabel 2 (zie p. 74) toont de itemhiërarchie zoals eerder vastgesteld met Mokken-schaalanalyse (Diesfeldt, 2012). De percentages juiste antwoorden uit het eerdere onderzoek zijn geactualiseerd met de gegevens uit het nieuwe onderzoek bij 65 patiënten (zie Box 2, p. 74).

Afgezien van kleine veranderingen in de percentages juiste antwoorden en het 95%-geloofwaardig interval blijft de moeilijkheidshiërarchie na opwaardering van de data met nieuwe gegevens gelijk. Deze moeilijkheidshiërarchie geldt op groepsniveau. Op individueel niveau kan de moeilijkheidshiërarchie echter variëren. De praktische consequenties daarvan worden zichtbaar in Figuur 1. De stringente procedure (CST9a: ‘stop na het eerste onjuist beantwoorde item’) leverde voor veel deelnemers een onderschatting op van de CST9-somscore. Tien deelnemers die volgens de stringente procedure een score 2 behaalden (geen juist antwoord bij item 3: ‘hoe heet de straat waar u nu woont?’), gaven bij items met een hogere moeilijkheidsgraad nog wel juiste antwoorden. Negen deelnemers konden de naam van de vorige koningin niet zeggen (item 5). Dat betekende een score 4 op de CST9a, maar acht van deze negen deelnemers behaalden op de CST9 een hogere score. Een dergelijke onderschatting maakt een strikt hiërarchische toepassing van de CST, ofwel stoppen na het eerste onjuist beantwoorde item, onwenselijk.



FIGUUR 1 De verhouding tussen scores volgens de hiërarchische stopregel (‘stop na het eerste onjuist beantwoorde item’; CST9a) en de compleet afgenomen hiërarchische variant (CST9). N = 65. Een cirkel vertegenwoordigt één of meer deelnemers.

TABEL 2 De opklimmende graad van moeilijkheid van negen CST-items zoals met Mokken-schaalanalyse vastgesteld (Diesfeldt, 2012; N = 642) en bayesiaans geactualiseerd met de gegevens van 65 nieuwe deelnemers: % correct en 95%-geloofwaardig interval (CI)

Item	Inhoud	Msa 2012 % correct (95%-CI)	Uit CST20 % correct (95%-CI)	CST9 % correct (95%-CI)
1	Geboortedatum	94,2 (92,3; 95,9)	94,5 (92,6; 96,0)	94,6 (92,8; 96,1)
2	Woonplaats	93,1 (91,0; 94,9)	93,0 (91,0; 94,8)	93,2 (91,1; 94,9)
3	Straat	87,2 (84,5; 89,7)	86,4 (83,7; 88,8)	86,4 (83,7; 88,8)
4	Koning/koningin	80,5 (77,4; 83,5)	80,5 (77,4; 83,3)	80,9 (77,9; 83,7)
5	Vorige koningin	71,2 (67,6; 74,6)	70,6 (67,2; 73,9)	71,2 (67,7; 74,4)
6	Maand	65,1 (61,4; 68,7)	66,2 (62,6; 69,6)	66,2 (62,6; 69,6)
7	Jaar	57,0 (53,2; 60,8)	57,7 (54,0; 61,3)	58,0 (54,3; 61,6)
8	Minister-president	42,4 (38,6; 46,2)	43,7 (40,1; 47,4)	44,2 (40,5; 47,8)
9	Vorige minister-president	16,0 (13,3; 19,0)	17,5 (14,8; 20,4)	17,9 (15,2; 20,9)

Afkortingen: Msa = Mokken-schaalanalyse; CST = Cognitieve Screening Test; 95%-CI = 95%-geloofwaardig (*credible*) interval op basis van bayesiaanse analyse

### Box 2 Bayesiaanse analyses

Voor de bepaling van de moeilijkheidsgraad van de items in de CST9 zijn de percentages juiste antwoorden uit eerder onderzoek in een bayesiaanse analyse geactualiseerd met de gegevens uit het nieuwe onderzoek bij 65 patiënten. In bayesiaanse termen wordt eerder verkregen informatie 'prior' genoemd, de informatie die een nieuwe studie oplevert heet 'likelijkheid' (Woertman e.a., 2014). 'Prior' en 'likelijkheid' zijn kansverdelingen die de waarschijnlijkheid van bepaalde waarden uitdrukken, zoals die van het percentage juiste antwoorden op CST-items, in een eerste, respectievelijk vervolgonderzoek. Door de 'prior' informatie uit eerder onderzoek te combineren met de 'likelijkheid' uit vervolgonderzoek wordt een 'posterior' berekend, ofwel de nieuwe kansverdeling van het percentage juiste antwoorden. De kansverdeling heeft een modale waarde en een spreiding. De modale waarde is het percentage juiste antwoorden, de spreiding wordt weergegeven door een 95%-geloofwaardig interval

(Kruschke, 2015). Het ‘geloofwaardig’ (credible) interval is een specifiek begrip in de bayesiaanse statistiek dat, anders dan het ‘betrouwbaarheidsinterval’ (confidence interval), de interpretatie toelaat dat het percentage juiste antwoorden zeer waarschijnlijk (95%) niet kleiner of groter is dan de laagste, respectievelijk hoogste waarde van het interval.

Ter toelichting het volgende voorbeeld. In de eerdere studie met 642 patiënten gaf 57% een juist antwoord op de vraag: welk jaar is het? Met een waarschijnlijkheid van 95% is het percentage juiste antwoorden niet kleiner dan 53,2% en niet groter dan 60,8% (zie Tabel 2, item 7). Na toevoeging van de gegevens uit de nieuwe studie wordt het percentage juiste antwoorden 57,7%, met een 95%-geloofwaardig interval van 54% tot 61,3%. De actualisering is twee keer berekend, op basis van de antwoorden ontleend aan identieke items in de CST20 en de antwoorden ontleend aan afname van de korte, hiërarchische CST9.

Ook correlaties en verschillen tussen correlaties (Tabel 3, p. 76) zijn in dit onderzoek door middel van bayesiaanse analyse bepaald en voorzien van een 95%-geloofwaardig interval voor de meest aannemelijke waarden van de correlaties of correlatieverschillen in het databestand (Kruschke, 2017).

### *Constructvaliditeit*

Tabel 3 laat zien dat alle versies van de CST positief correleren met de VAT. Bij toetsing van de verschillen tussen de correlaties kon in deze relatief kleine steekproef geen enkele nulhypothese (‘de correlaties verschillen niet’) worden verworpen. Een nulhypothese niet kunnen verwerpen geeft op zichzelf geen inzicht in de grootte van de verschillen tussen de onderzochte correlaties (Lakens e.a., 2020). Bayesiaanse analyse maakt de grootte van de verschillen juist wel inzichtelijk. Het grootste deel (92%) van de posterior kansverdeling voor het verschil tussen de twee correlaties VAT met CST20 en VAT met CST9 is positief (de correlatie VAT met CST20 is groter), maar een negatief verschil (de correlatie VAT met CST20 is kleiner) is met 8% (100% - 92%) niet uitgesloten. Ten opzichte van de VAT zijn CST9 en CST9a even valide, met correlaties van respectievelijk 0,60 en 0,58. Met een verschil van 0,02 ontlopen de correlatiecoëfficiënten elkaar niet veel (63% van de posterior kansverdeling > 0; 37% < 0). Gelet op het 95%-geloofwaardig interval zijn een negatief verschil van -0,09 of een positief verschil van 0,12 echter niet uitgesloten. Het interval is ruim vanwege de relatief kleine steekproef.



TABEL 3 Gemiddelden, standaarddeviaties en correlaties tussen VAT, CST20, CST9 en CST9a; Pearsons r met bayesiaans 95%-geloofwaardig interval, N = 55

	M (SD)	CST20	CST9	CST9a
VAT	7,3 (4,4)	0,67 (0,49; 0,79)	0,60 (0,39; 0,74)	0,58 (0,37; 0,73)
CST20	14,6 (4,1)	1	0,88 (0,80; 0,93)	0,81 (0,69; 0,88)
CST9	6,7 (2,3)	0,07 (-0,03; 0,19)	1	0,90 (0,83; 0,94)
CST9a	5,7 (2,9)	0,08 (-0,04; 0,23)	0,02 (-0,09; 0,12)	1

Afkortingen: M = gemiddelde, SD = standaarddeviatie; CST = Cognitieve Screening Test; VAT = Visuele Associatie Test. Legenda: boven de diagonaal (zie de cijfers 1) zes correlatiecoëfficiënten (met bayesiaans 95%-geloofwaardig interval); onder de diagonaal drie correlatieverschillen (met bayesiaans 95%-geloofwaardig interval): de correlatie tussen VAT en CST9 vergeleken met de correlatie tussen VAT en CST20, de correlatie tussen VAT en CST9a vergeleken met de correlatie tussen VAT en CST20 en de correlatie tussen VAT en CST9a vergeleken met de correlatie tussen VAT en CST20

### Test-hertesteffecten

De CST9-resultaten waren bij tweede afname iets beter dan bij eerste afname, met een gemiddeld verschil van 0,38 (zie Tabel 4). Het maakte daarbij geen verschil of de negen items bij eerste afname in de context van de CST20 waren aangeboden of in de hiërarchische volgorde, als op zichzelf staande CST9.

Tabel 4 laat een test-hertestbetrouwbaarheid van 0,89 zien (berekend als Intraclass Correlatiecoëfficiënt, ICC). Voor toepassing in de klinische praktijk is de 95%-Reliable Change Index (RCI) van belang. De RCI is het klinisch relevante verschil tussen een test- en een hertestscore. Voor de CST9 zijn een *scorevermindering* tussen twee testafnames van (afgerond) twee punten of meer, respectievelijk een *scoretoename* van drie punten of meer, betekenisvol. Kleinere scoreveranderingen kunnen beschouwd worden als normale fluctuaties in herhalingsonderzoek (Woods e.a., 2006).

TABEL 4 Test-hertestgegevens (N = 65) voor de hiërarchische variant van de Cognitieve Screening Test: CST9

Test	Hertest	T	p (tweezijdig)	ICC	Verskil hertest-test	
					M (SD)	95%-RCI
6,37 (2,23)	6,75 (2,34)	3,02	0,004	0,89	0,38 (1,03)	-1,63; 2,40

M = gemiddelde; SD = standaarddeviatie; ICC = Intraclass Coëfficiënt; RCI = Reliable Change Index

Leereffecten tussen testafnames betroffen vaker de semantische dan de episodische geheugenitems. Tabel 5 laat een verbetering tussen test en hertest zien voor de gemiddelde somscore van de zeven semantische geheugenitems, maar niet voor die van de twee episodische geheugenitems: 'wat is de maand' en 'wat is het jaar'. Het verschil in leereffect voor semantische en episodische geheugenitems komt tot uitdrukking in de significante interactieterm: test-hertest met itemtype (episodisch of semantisch), in een variantieanalyse met herhaalde metingen. Episodische geheugenitems vragen telkens weer om inzicht in een veranderlijke temporele context. Semantische geheugenitems daarentegen doen een beroep op verworven kennis, die in of na een eerste testsessie kan zijn geactiveerd en in een volgende testsessie opnieuw of beter toegankelijk is. Zo werd een sterk netto leereffect gevonden voor item 8, waarbij acht van de 65 deelnemers premier Mark Rutte niet bij de eerste testsessie, maar wel bij de tweede wisten te noemen, tegenover één deelnemer voor wie het omgekeerde gold.

Er werd geen verband gevonden tussen de grootte van een leereffect en een langer of korter interval tussen de twee testsessies. De plausibele verwachting van een negatieve correlatie, namelijk dat een korter interval zou samenhangen met een groter leereffect, werd niet door de data ondersteund ( $r = 0,09$ ; 95%-CI:  $-0,16$ ;  $0,31$ ).

TABEL 5 Test-hertestgegevens (N = 65) onderscheiden naar de somscores voor de twee episodische en de zeven semantische geheugenitems in de hiërarchische variant van de Cognitieve Screening Test: CST9

CST9	Test	Hertest
	M (SD)	M (SD)
Episodische geheugenitems	1,45 (0,73)	1,41 (0,77)
Semantische geheugenitems	4,92 (1,68)	5,34 (1,74)

M = gemiddelde; SD = standaarddeviatie; F-toets voor de interactieterm geheugentype\*testmoment:  $F_{1,64} = 11,49$ ,  $p = 0,001$

## Discussie

De hiërarchische versie van de Cognitieve Screening Test is hier op bruikbaarheid voor de klinische praktijk onderzocht. Er zijn twee manieren om een hiërarchisch gestructureerde schaal te gebruiken: stoppen na het

eerste item dat onjuist wordt beantwoord of alle items afnemen. De oorspronkelijk vastgestelde opklimmende moeilijkheidsgraad van negen CST-items werd bevestigd, maar is niet voor iedereen even stringent. Het advies is dan ook om bij klinische toepassing alle negen items voor te leggen (Kempen e.a., 1995). Overigens verschilden de validiteitscoëfficiënten (correlaties met een andere geheugentest (VAT)) bij toepassing van de stopregel of volledige afname weinig.

Wat betekent de CST9 voor de klinische praktijk? De CST9 is een optimale selectie van gemakkelijke en moeilijke items uit de CST20. De negen items representeren de essentie van de oorspronkelijke 20-itemversie en vormen een bruikbaar instrument voor onderzoek van de mentale status. Test-hertestgegevens bieden een referentie die de klinisch neuropsycholoog kan gebruiken om betekenisvolle veranderingen vast te stellen, zoals een scoreverbetering van minstens drie punten of een scorevermindering van twee punten of meer. Deze gegevens zijn voor de CST20 niet beschikbaar. Voor onderzoek van (oudere) patiënten die vanwege verminderde cognitieve capaciteiten weinig belastbaar zijn, kan gebruik van de korte, hiërarchische CST9 de voorkeur verdienen boven de langere CST20.

### Dankbetuiging

De auteur is de psychologen Judith Centen, Nicole Hooiveld, Mary Koer en Anneke Schoen zeer erkentelijk voor hun medewerking aan dit onderzoek.

**Han Diesfeldt** Zelfstandig onderzoeker, Castricum, e-mail: h.diesfeldt@outlook.com.

#### Literatuur

- Diesfeldt, H. (2003). De minister-president in het onderzoek van de mentale status. *Tijdschrift voor Gerontologie en Geriatrie*, 34, 168-175.
- Diesfeldt, H. (2006). Constructvaliditeit van enkele tests voor episodisch geheugen in de psychogeriatric. *Tijdschrift voor Gerontologie en Geriatrie*, 37, 59-66.
- Diesfeldt, H. (2012). De Cognitieve Screening Test (CST) bij deelnemers aan psychogeriatric dagbehandeling: Voorstel voor een hiërarchische schaal. *Tijdschrift voor Neuropsychologie*, 7, 90-98.
- Graaf, A. de & Deelman, B. G. (1991). *Cognitieve Screening Test*. Lisse: Swets & Zeitlinger.
- Guilmette, T.J., Sweet, J.J., Hebben, N., Kol-tai, D., Mahone, E.M., Spiegler, B.J. e.a. (2020). American Academy of Clinical Neuropsychology consensus conference statement on uniform labeling of performance test scores. *The Clinical Neuropsychologist*, 34, 437-453.
- Hendriks, M.P.H., Mol, B.A.W. & Kessels, R.P.C. (2020). Opinie: Uniformiteit in de kwalitatieve beschrijving van scores op prestatietaken. *Tijdschrift voor Neuropsychologie*, 15, 166-176.

- Kempen, G.I.J.M., Myers, A.M. & Powell, L.E. (1995). Hierarchical structure in ADL and IADL: Analytical assumptions and applications for clinicians and researchers. *Journal of Clinical Epidemiology*, 48, 1299-1305.
- Kruschke, J.K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Londen: Academic Press.
- Kruschke, J.K. (red.), (2017). *Bayesian estimation of correlations and differences of correlations within a multivariate normal*. Doingbayesiandataanalysis.blogspot.
- Lakens, D., McLatchie, N., Isager, P.M., Scheel, A.M. & Dienes, Z. (2020). Improving inferences about null effects with Bayes factors and equivalence tests. *Journal of Gerontology: Series B*, 75, 45-57.
- McGrory, S., Austin, E.J., Shenkin, S.D., Starr, J.M. & Deary, I.J. (2015). From 'aisle' to 'labile': A hierarchical National Adult Reading Test scale revealed by Mokken scaling. *Psychological Assessment*, 27, 932-943.
- NIP. (2010). *Richtlijn voor een kort neuropsychologisch onderzoek bij patiënten met een beroerte*. Utrecht: NIP: sectie Revalidatie en sectie Neuropsychologie.
- Pastink, C., Schepel, C., Kingma, M. & Rietkerk, I. (2019). *Neuropsychologische diagnostiek bij ouderen*. Leeuwarden: Ouderpsychologie Noord.
- Sijtsma, K. & Van der Ark, L.A. (2017). A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *British Journal of Mathematical and Statistical Psychology*, 70, 137-158.
- Stochl, J., Jones, P.B. & Croudace, T.J. (2012). Mokken scale analysis of mental health and wellbeing questionnaire item responses: A nonparametric IRT method in empirical research for applied health researchers. *BMC Medical Research Methodology*, 12(74), 1-16.
- Streiner, D.L., Norman, G.R. & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use. Fifth edition*. Oxford: Oxford University Press.
- Toutert, M. van, Diesfeldt, H. & Hoek, D. (2016). De Cognitieve Screening Test (CST) bij normale cognitieve veroudering en dementie: Drie varianten en bruikbaarheid voor de klinische praktijk. *Tijdschrift voor Neuropsychologie*, 11, 274-292.
- Verhage, F. (1964). *Intelligentie en leeftijd bij volwassenen en bejaarden*. Assen: Van Gorcum.
- Watson, R., Deary, I.J. & Shipley, B. (2008). A hierarchy of distress: Mokken scaling of the GHQ-30. *Psychological Medicine*, 38, 575-579.
- Watson, R., Van der Ark, L.A., Lin, L., Fieo, R., Deary, I.J. & Meijer, R.R. (2012). Item response theory: How Mokken scaling can be used in clinical practice. *Journal of Clinical Nursing*, 19-20, 2736-2746.
- Wattel, L. & Achterberg, W. (2010). *UNOCOG toolkit: Cognitieve meetinstrumenten voor de ouderenzorg: Een practice based benadering*. Amsterdam: Universitair Netwerk Ouderenzorg UNO-VUmc.
- Woertman, W.H., Groenewoud, H.M.M. & Van der Wilt, G.J. (2014). Bayesiaanse statistiek. Wat, hoe en waarom? *Nederlands Tijdschrift voor Geneeskunde*, 158(A7485), 1-6.
- Woods, S.P., Delis, D.C., Scott, J.C., Kramer, J.H. & Holdnack, J.A. (2006). The California Verbal Learning Test-second edition: Test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms. *Archives of Clinical Neuropsychology*, 21, 413-420.