

De Cognitieve Screening Test (CST) getoetst aan het schaalmodel van Mokken

H.F.A. Diesfeldt^a

Mokken scaling of the Cognitive Screening Test

The Cognitive Screening Test (CST) is a twenty-item orientation questionnaire in Dutch, that is commonly used to evaluate cognitive impairment. This study applied Mokken Scale Analysis, a non-parametric set of techniques derived from item response theory (IRT), to CST-data of 466 consecutive participants in psychogeriatric day care. The full item set and the standard short version of fourteen items both met the assumptions of the monotone homogeneity model, with scalability coefficient $H = 0.39$, which is considered weak. In order to select items that would fulfil the assumption of invariant item ordering or the double monotonicity model, the subjects were randomly partitioned into a training set (50% of the sample) and a test set (the remaining half). By means of an automated item selection eleven items were found to measure one latent trait, with $H = 0.67$ and item H coefficients larger than 0.51. Cross-validation of the item analysis in the remaining half of the subjects gave comparable values ($H = 0.66$; item H coefficients larger than 0.56). The selected items involve year, place of residence, birth date, the monarch's and prime minister's names, and their predecessors. Applying optimal discriminant analysis (ODA) it was found that the full set of twenty CST items performed best in distinguishing two predefined groups of patients of lower or higher cognitive ability, as established by an independent criterion derived from the Amsterdam Dementia Screening Test. The chance corrected predictive value or prognostic utility was 47.5% for the full item set, 45.2% for the fourteen items of the standard short version of the CST, and 46.1% for the homogeneous, unidimensional set of selected eleven items. The results of the item analysis support the application of the CST in cognitive assessment, and revealed a more reliable 'short' version of the CST than the standard short version (CST14).

Keywords: Non-parametric item response theory, optimal discriminant analysis, orientation questionnaire, psychogeriatric day care, dementia
Tijdschr Gerontol Geriatr 2009; 40: 193-202

^a PgD, De Stichtse Hof, Vivium zorggroep, Laren NH
Correspondentie: Dr. H.F.A. Diesfeldt, PgD, De Stichtse

Hof, Naarderstraat 81, 1251 BG Laren. E:
h.diesfeldt@vivium.nl

Samenvatting

De itemreeks van de Cognitieve Screening Test (CST) werd getoetst aan een niet-parametrisch model uit de itemresponsetheorie (IRT), namelijk het schaalmodel van Mokken. Hiervoor waren gegevens beschikbaar van 466 opeenvolgende deelnemers aan een psychogeriatrische dagbehandeling. De volledige CST (20 items) en de standaard verkorte, 14-item versie van de CST voldeden aan de eerste monotonie-eis van het schaalmodel van Mokken, met in beide gevallen een zwakke schaalbaarheidscoëfficiënt H van 0,39. Voor de bepaling van een optimale itemverzameling, die voldeed aan de voorwaarde van unidimensionaliteit (dubbele monotonie), werd at random een 'training set' van 233 deelnemers getrokken. Automatische selectie bracht een verzameling van elf items aan het licht met een H -waarde van 0,67 en item- H -waarden van 0,52 of hoger. Deze items werden door kruisvalidering in een tweede deelsteekproef (de 'test set'; $n=233$) op hun schaalbaarheid getoetst ($H=0,66$; item- H -waarden 0,57 of hoger). De geselecteerde items betreffen het huidige jaar, de woonplaats, de geboortedatum van de onderzochte, en de namen van vorstin en premier, inclusief die van hun ambtsvoorgangers. Door middel van non-parametrische optimale discriminantanalyse werd het onderscheidend vermogen van diverse CST-itemcombinaties onderzocht met de Amsterdamse Dementiescreeningstest (ADS) als onafhankelijk criterium. Gelet op het aantal juiste classificaties had de CST₂₀ het hoogste onderscheidend vermogen, met een voor toeval gecorrigeerde voorspellende waarde (prognostic utility) van 47,5%. Voor de CST₁₄ was de prognostic utility 45,2% en voor de selectie van elf CST-items: 46,1%. Gelet op schaalbaarheid en validiteit is de hier gevonden selectie van elf items een nauwkeuriger korte versie van de CST dan de standaard korte versie 'CST₁₄'.

Trefwoorden: Cognitieve Screening Test (CST); Amsterdamse Dementiescreeningstest (ADS); Mokkens schaalmodel; psychogeriatrische dagbehandeling

Inleiding

In 1991 kwam voor klinisch gebruik in de psychogeriatric de Cognitieve Screening Test (CST) beschikbaar. De CST heeft twintig items en bevat oriëntatievragen, vragen naar bekende publieke personen (zoals de huidige koningin, haar voor-

gangster, de minister-president) en vragen naar feiten uit voorbije jaren (zoals de jaartallen van de twee wereldoorlogen).¹ Voor toepassing in de klinische praktijk is het belangrijk dat de somscore over de items van de CST betekenisvolle uitspraken toelaat over een eigenschap van de respondenten, zoals in dit geval hun niveau van cognitief functioneren. Betekenisvol kan op minstens twee manieren worden opgevat, in de zin van betrouwbaarheid en validiteit. Om betrouwbaar individuele verschillen te kunnen vaststellen, is het van belang dat de testscore respondenten systematisch ordent op de gemeten eigenschap.² Validiteit heeft betrekking op het verband tussen de gemeten (latente) eigenschap en de resultaten van andere metingen of observaties in verwante domeinen.

Het vermogen van de CST om respondenten betrouwbaar te ordenen op de latente eigenschap 'niveau van cognitief functioneren' is hier onderzocht met behulp van een niet-parametrisch model uit de itemresponsetheorie (IRT), namelijk het schaalmodel van Mokken. IRT-modellen brengen het antwoordgedrag van personen op een reeks vragen in kaart.³ Het schaalmodel van Mokken is geschikt voor onderzoek en samenstelling van itemverzamelingen die in de praktijk een betrouwbare rangordening van respondenten mogelijk moeten maken.⁴ Het niet-parametrische model van Mokken stelt als voorwaarde aan een passend item dat de kans om op dat item een juist antwoord te geven monotoon toeneemt met de (latente) vaardigheid die dat item beoogt te meten. Parametrische IRT-modellen stellen daarentegen extra eisen aan het kansverloop door hiervoor een specifieke wiskundige vorm voorop te stellen die aan de hand van een beperkt aantal parameters kan worden beschreven. Items waarvan de observaties niet overeenkomen met het voorgestelde kansverloop, kunnen niet worden gemodelleerd en worden uit de schaal verwijderd, ook al kunnen ze inhoudelijk interessant zijn. Niet-parametrische modellen hebben dit nadeel niet, en zijn geschikt om itemverzamelingen te vinden waarmee respondenten betrouwbaar kunnen worden geordend op de gemeten eigenschap.

Eerste doel van het hier te beschrijven onderzoek was de complete itemreeks van de CST en de veertien items van de standaard verkorte versie van de CST te toetsen aan het schaalmodel van Mokken. Dit werd gedaan op basis van responsies van een grote groep psychogeriatrische patiënten. De betrouwbaarheid van de itemreeks werd aan twee belangrijke theoretische eigenschappen van het model getoetst. Ten eerste werd nagegaan of de items betrekking hebben op hetzelfde continuüm, in dit geval: niveau van cognitief functioneren. De veronderstelling dat de items eendimensionaal of homogeen zijn, houdt volgens het schaalmodel in dat de onderzochte deelnemers, gegeven hun niveau van cognitief functioneren, door elk item op dezelfde wijze worden geordend.⁵ Volgens de tweede eigenschap van het model zijn ook de items, gegeven hun moeilijkheidsgraad, voor elke

respondent op dezelfde wijze geordend. Dat wil zeggen: wanneer één item moeilijker is dan een ander item, dan geldt dat altijd, ongeacht het niveau van cognitief functioneren van de respondent. Wanneer dit empirisch kan worden aangetoond, voldoet de vragenlijst of test aan de voorwaarde van dubbele monotonie. Elk item houdt zich dan steeds op dezelfde manier tot de te meten eigenschap. Voor de praktijk is dat van belang, omdat daarmee wordt aangetoond dat de score op een item op dezelfde wijze bijdraagt aan de schatting van iemands positie op het continuüm van de gemeten eigenschap, ongeacht of de respondent bijvoorbeeld een man of een vrouw is, of een hoog of een laag opleidingsniveau heeft.

Wanneer een standaard itemverzameling, zoals de CST, niet of onvolledig voldoet aan het Mokkenmodel, kan worden gezocht naar een itemcombinatie die dat wel doet. Dit was het tweede doel van het onderzoek: een itemreeks selecteren die aan hoge eisen van schaalbaarheid voldoet.

Het derde onderzoeksdoel was om de validiteit van de verschillende itemverzamelingen te onderzoeken. Dit werd gedaan door de correlatie te berekenen tussen de somscores van de verschillende itemverzamelingen en de resultaten van een andere, meer uitgebreide indicator van cognitief functioneren, namelijk de Amsterdamse Dementiescreeningstest (ADS).

Methode

Deelnemers

Deelnemers aan dit onderzoek waren alle 466 patiënten die in 2003-2008 een psychogeriatrische dagbehandeling bezochten. Zij werden voor dagbehandeling verwezen door geriateren of neurologen van de regionale geheugenpoli's. Enkele demografische gegevens van deze groep zijn weergegeven in Tabel 1.

| Tabel 1 | Demografische kenmerken. |
|--|--------------------------|
| Variabele | N = 466 |
| Geslacht (%) | |
| • man | 41,0 |
| • vrouw | 59,0 |
| Leeftijd (gemiddeld; SD) | 78,9 (6,3) |
| Opleiding (%) ^a | |
| • lager onderwijs | 20,0 |
| • lager beroepsonderwijs | 32,6 |
| • ulo of mavo | 20,8 |
| • havo of vwo | 21,7 |
| • universiteit | 4,9 |
| Woonsituatie (%) | |
| • alleen | 47,6 |
| • samen | 52,4 |
| Diagnose (%) | |
| • dementie | 95,1 |
| • anders | 4,9 |
| Duur dementie in jaren (gemiddeld; SD) | 3,5 (3,0) ^b |
| ADS ₃ (gemiddeld; SD) | -0,3 (2,7) ^c |

^a Volgens de classificatie van Verhage²¹

^b n = 443

^c n = 456

Diagnosegegevens werden gecodeerd volgens de systematiek van de DSM-IV, op basis van gegevens van de verwijzende specialist.⁶ De duur van een eventuele dementie werd bepaald op grond van informatie uit het dossier. Bij 95% van de deelnemers werd een vorm van dementie vastgesteld, met een gemiddelde duur van 3,5 jaar ($SD=3,0$). Nosologische oorzaken waren ziekte van Alzheimer (72%), vasculaire dementie (17%) en overige dementieën (6%), bijvoorbeeld als gevolg van de ziekte van Parkinson of frontotemporale atrofie. Bij 5% van de patiënten was geen dementie gevonden, maar een affectieve stoornis of een enkelvoudige cognitieve stoornis. De deelnemers zijn te kenschetsen als patiënten met een cognitieve stoornis, variërend van een lichte tot matig ernstige dementie, die met behulp van mantelzorg thuis woonden.

Werkwijze

In het kader van de dagbehandeling werd elke patiënt uitgenodigd voor een psychologisch onderzoek, volgens een vast protocol. Het onderzoek was gespreid over twee sessies met een interval van meestal een week. In de eerste sessie werd de CST afgenomen, in de tweede de ADS. Aan de ADS werd een indicator voor het globale niveau van cognitief functioneren ontleend in de vorm van de somscore over drie subtests die met elkaar de ADS₃, de standaard korte versie van de ADS vormen.⁷ De ADS₃ bestaat uit het onthouden van afbeeldingen (*Visueel Geheugen*), tijd- en plaatsoriëntatie, met drie vragen naar maand, jaar en naam van de instelling (*Oriëntatie*), en opnoemen van zoveel mogelijk dieren en beroepen, telkens in één minuut (*Fluency*). Gewogen scores per subtest vormen een ordinale schaal van -2 tot +1 (voor *Visueel Geheugen* en *Oriëntatie*), en van -2 tot +2 voor *Fluency*. Lage gewogen scores passen met een grotere waarschijnlijkheid bij dementie dan hoge gewogen scores. De som van de gewogen scores over de drie subtests varieert van -6 tot +4, waarbij hoge scores een relatief hoog niveau van cognitief functioneren weergeven. Tien deelnemers waren voor de geplande tweede zitting, waarin de ADS zou worden afgenomen, niet meer in dagbehandeling. Daardoor komt het totale aantal patiënten voor wie gegevens van de CST konden worden vergeleken met de ADS op 456.

Analyses

Voor de schaalanalyse van de CST is de Mokken Schaalanalyse voor Polytome items (MSP) gebruikt.⁸ MSP is zowel voor dichotoom gescoorde items als voor polytome items (meer dan twee antwoordcategorieën) geschikt.⁴ De CST heeft dertien dichotoom gescoorde items (twee antwoordcategorieën). Zeven items worden poly-

toom gescoord (0, 0,5 en 1) om recht te doen aan een gedeeltelijk juist antwoord, zoals een maand of dag van de week die direct voorafgaat of volgt op de eigenlijke maand of weekdag, of enkel het begin- of eindjaar van een wereldoorlog. Voor de analyse werden dichotome items gecodeerd als 0 en 1 (voor een onjuist, respectievelijk juist antwoord). Het programma MSP accepteert geen itemscores met een decimaal, wat werd opgelost door de polytome items 0, 0,5 en 1 te coderen als respectievelijk 0, 1 en 2.⁹ Scoring gebeurde volgens de handleiding van de CST.¹ Over scoring van het antwoord op de vraag naar het jaargetijde is de handleiding niet specifiek. In dit onderzoek is het antwoord 'lente' of 'voorjaar' goed gerekend wanneer dit werd gegeven in de maanden maart, april, mei en juni. Het antwoord 'zomer' werd goed gerekend in de maanden juni tot en met september, het antwoord 'herfst' of 'najaar' in de maanden september tot en met december, en het antwoord 'winter' in de maanden december tot en met maart.

Het programma MSP berekent een schaalbaarheidscoëfficiënt H die de waarschijnlijkheid aangeeft waarmee de itemverzameling beantwoordt aan een 'perfecte' (hiërarchische) Guttman-schaal.¹⁰ De coëfficiënt van schaalbaarheid (of homogeniteit) wordt niet alleen voor de schaal in zijn geheel (H), maar ook per item berekend (H_i). Wanneer H en H_i groter zijn dan 0,50, wijst dit op een sterke schaal. Coëfficiënten tussen 0,40 en 0,50 wijzen op een matig sterke schaal, en coëfficiënten tussen 0,30 en 0,40 op een zwakke schaal. Items met een H-coëfficiënt < 0,30 zijn slecht schaalbaar. Zij dragen niet betekenisvol bij aan de somscore of indicator van de te meten eigenschap.⁴

MSP kent een algoritme dat op basis van interitemcorrelaties en item-H-coëfficiënten zoekt naar de combinatie van twee items met de hoogste H-waarde. Daar wordt vervolgens een ander item bij gezocht, zodanig dat H voor de nieuwe combinatie opnieuw zo hoog mogelijk is. Dit automatische zoekproces gaat door totdat een door de onderzoeker gekozen ondergrens van H wordt bereikt. Om een itemverzameling te vinden die beantwoordt aan de eisen van een 'sterke' schaal, is als ondergrens een H-waarde van 0,51 genomen. Deze functionaliteit van MSP is toegepast op de twintig items van de CST met als doel een itemverzameling te vinden die past bij een IRT-model van dubbele monotonie.

Om de generaliseerbaarheid van de automatische itemselectie te vergroten werd de steekproef gesplitst. Dit gebeurde op willekeurige wijze met behulp van de opdracht 'aselect' in het rekenprogramma Excel. De helft van de steekproef (de 'training set'; $n=233$) werd gebruikt voor de automatische itemselectie. Vervolgens werd kruisvalidering toegepast in de tweede deelsteekproef (de 'test set') om de passing van de geselecteerde items te toetsen.

| Tabel 2 Items van de Cognitieve Screening Test (CST): gemiddelde (M), standaarddeviatie (SD) en item-schaalbaarheidscoëfficiënt H_i . Voor de volledige schaal (CST20) en de standaard korte versie (CST14): gemiddelde (M), standaarddeviatie (SD), schaalbaarheidscoëfficiënt H en coëfficiënt voor interne consistentie (Rho). N=466. | | | | | |
|---|--|----|-------------|---------------------|---------------------|
| Item | Tekst | AC | M (SD) | CST20 H per item | CST14 H per item |
| 1 | Welk jaar is het? | 2 | 0,58 (0,49) | 0,46 | 0,47 |
| 2 | Welke maand is het? | 3 | 0,74 (0,40) | 0,41 | 0,42 |
| 3 | De hoeveelste van de maand is het? | 2 | 0,50 (0,50) | 0,41 | 0,48 |
| 4 | Welke dag van de week is het? | 3 | 0,78 (0,39) | 0,37 | 0,37 |
| 5 | Hoe heet de plaats waar u woont? | 2 | 0,92 (0,26) | 0,50 | 0,49 |
| 6 | Hoe heet de straat waar u woont? | 3 | 0,88 (0,33) | 0,40 | 0,38 |
| 7 | Hoe oud bent u? | 3 | 0,70 (0,42) | 0,30 | 0,34 |
| 8 | In welk jaar bent u geboren? | 2 | 0,95 (0,23) | 0,54 | 0,50 |
| 9 | In welke maand bent u geboren? | 2 | 0,98 (0,14) | 0,78 | 0,76 |
| 10 | Op de hoeveelste van de maand bent u geboren? | 2 | 0,97 (0,16) | 0,66 | 0,66 |
| 11 | Wie is op het ogenblik onze koningin? | 2 | 0,80 (0,40) | 0,44 | 0,42 |
| 12 | Wie was koningin voor haar? | 2 | 0,71 (0,45) | 0,43 | 0,44 |
| 13 | (Volgorde juist?) | 2 | 0,74 (0,44) | 0,40 | 0,39 |
| 14 | Hoe laat is het nu (zonder op horloge of klok te kijken) | 3 | 0,58 (0,43) | 0,22 | 0,24 |
| 15 | Wie is op het ogenblik onze minister-president? | 2 | 0,45 (0,50) | 0,44 | . |
| 16 | Wie was minister-president voor hem? | 2 | 0,19 (0,39) | 0,58 | . |
| 17 | (Volgorde juist?) | 2 | 0,16 (0,36) | 0,65 | . |
| 18 | Wanneer was de Eerste Wereldoorlog? | 3 | 0,62 (0,46) | 0,30 | . |
| 19 | Wanneer was de Tweede Wereldoorlog? | 3 | 0,70 (0,44) | 0,41 | . |
| 20 | Welk jaargetijde is het? | 2 | 0,66 (0,47) | 0,34 | . |
| | CST20: M (SD) | | 13,6 (4,2) | | |
| | CST14: M (SD) | | 10,9 (2,9) | | |
| | Mokkens H | | | 0,39 | 0,39 |
| | NIRT-model | | | MHM | MHM |
| | Rho | | | 0,86 | 0,82 |

AC = aantal antwoordcategorieën: 2 (voor de itemscores 0 of 1), of 3 (voor de itemscores 0, 0,5 of 1); NIRT-model = niet-parametrisch item-responsetheoriemodel; MHM = monotoon homogeen model.

De validiteit van de CST werd getoetst door deze te vergelijken met de ADS₃. Voor deze vergelijking werden de deelnemers verdeeld over twee scoreniveaus van de ADS₃, een groep met een ADS-somscore van 0 of minder, en een groep met een somscore van 1 of meer. Deze cesuur is het kantelpunt waaronder of waarboven een dementie meer, respectievelijk minder waarschijnlijk is.¹¹ Met behulp van discriminantanalyse is vervolgens nagegaan in hoeverre de somscore op een itemverzameling van de CST een juiste verdeling mogelijk maakte van patiënten over de twee ADS-niveaus.

Voor de discriminantanalyse is een niet-parametrische techniek gebruikt met behulp van ODA-software (*Optimal Data Analysis*).¹² Het in ODA gebruikte algoritme vindt precies die grensscore die met maximale nauwkeurigheid onderscheid maakt tussen twee (of meer) vooraf gedefinieerde groepen, en zou dat volgens de ontwerpers van de software nauwkeuriger doen dan een parametrische discriminantanalyse of logistische regressieanalyse. De kwaliteit van de analyse wordt afgeleid uit diverse indicatoren, zoals het aantal correcte classificaties (CC), de sensitiviteit, de specificiteit en de *prognostic utility* (PU). De PU is de voor toeval gecorrigeerde, gemiddelde voorspellende waarde van een lage, respectievelijk hoge score op een CST-itemverzameling voor respectievelijk lage of hoge ADS-scores. De PU kan variëren van 0% tot 100%, waarbij 0% duidt op een onderscheidend vermogen dat niet uitstijgt boven kansniveau, en 100% een perfecte classificatie betekent. Een PU < 25% wijst op een zwak verband tussen twee variabelen, een PU tussen 25% en 50% op een matig sterk verband. Bij een PU > 50% is sprake van een 'robuust' verband.¹²

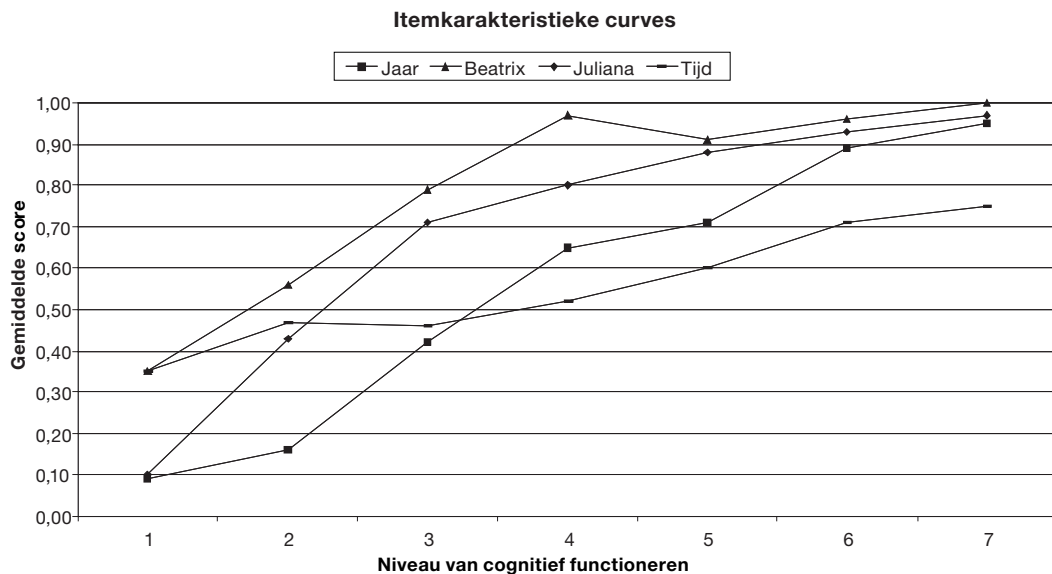
Resultaten

Het percentage juiste antwoorden op de items van de CST varieerde van 19% (voor 'Wim Kok' als ambtsvoorganger van premier Jan Peter Balkenende) tot 98% voor de eigen geboortemaand. Tabel 2 geeft deze percentages weer.

De gemiddelde score van 0,16 op item 17 (juiste opeenvolging van premiers) betekent dat 16% van de deelnemers de namen van beide premiers kon noemen. Iets meer dan de helft van de deelnemers (52,6%) wist geen antwoord op de vraag naar de naam van de minister-president (in de onderzochte periode was dat 'Balkenende'), en een kleine minderheid (2,6%) noemde een andere naam, zoals Kok, Lubbers, Van Agt, Zalm, Drees, Biesheuvel of Gerbrandy. Op de vraag naar de ambtsvoorganger van Balkenende wist 72% geen antwoord, en noemde 8,6% een andere naam dan Wim Kok, met als toppers Lubbers (14 keer) en Den Uyl (8 keer).

Meer dan 90% van de deelnemers wist de eigen woonplaats te noemen, of de geboortedatum. Bijna iedereen (99%) noemde minstens één onderdeel van de geboortedatum; 94% noemde de geboortedatum compleet. Meer deelnemers kenden de jaartallen van de Tweede Wereldoorlog dan van de Eerste. Ook valt op dat veel meer deelnemers de naam van de koningin wisten (80%) dan die van de premier (45%).

De homogeniteitscoëfficiënt H was 0,39 voor de complete itemverzameling, evenals voor de standaard korte, 14 item-versie van de CST. Beide schalen voldoen aan de eerste monotonie-eis van het Mokkenmodel, maar een H-waarde < 0,40 geldt als zwak. De H-waarde werd gedrukt door één item (item 14; hoe laat is het nu?) met een H_i-coëfficiënt van 0,22. De itemkarakteristieke



Figuur 1

Gemiddelde scores op vier items van de CST gerelateerd aan de hoogte van de restscore (niveau van cognitief functioneren) volgens zeven klassen (van laag naar hoog). N per klasse varieert van 51 tot 86.

curve van dit item is weergegeven in Figuur 1. De figuur laat zien dat de kans op een juist antwoord op de vraag naar de tijd (item 14) al tamelijk groot is (35%) voor mensen met een relatief laag niveau van cognitief functioneren (klasse 1 op de horizontale as). Vervolgens nemen de percentages slechts matig toe met een toenemend niveau van cognitief functioneren. De itemkarakteristieke curve van dit item verloopt nagenoeg vlak. Het item draagt daarom weinig bij aan het onderscheid tussen deelnemers met een relatief laag niveau van cognitief functioneren en deelnemers met een hoger niveau.

Figuur 1 laat ook de itemkarakteristieke functies zien van drie andere items met item H-waarden in het matig sterke gebied (0,40-0,50), zoals de vragen naar het huidige jaar, en de

naam van voormalig koningin Juliana, respectievelijk koningin Beatrix. De figuur laat zien hoe de gemiddelde score van deze items toeneemt met een toenemend niveau van cognitief functioneren, hier uitgedrukt in zeven klassen van de restscore, dat is de somscore over alle items behalve die van het item waarvan de itemkarakteristieke curve wordt bepaald. Ook is te zien hoe de drie itemkarakteristieke curves met hoge item-H-waarden elkaar niet snijden, en daarmee beantwoorden aan de voorwaarde van dubbele monotonie. Dit geldt niet voor item 14 (hoe laat is het nu?). Wanneer dit item uit de analyse wordt weggelaten, stijgt de H-coëfficiënt naar 0,42 (voor de overige negentien items van de CST20) of naar 0,44 (voor de overige dertien items van de CST14).

Tabel 3 Een meettheoretisch optimale selectie van elf CST-items voor een willekeurig gekozen 'training set' en een tweede deelsteekproef (de 'test set'). Gemiddelden (M), standaarddeviaties (SD), itemschaalbaarheidscoëfficiënten H_i , schaalbaarheidscoëfficiënten H en Rho.

| Item | Tekst | Training set (n = 233) | | Test set (n = 233) | |
|------|---|------------------------|------------|--------------------|------------|
| | | M (SD) | H per item | M (SD) | H per item |
| 1 | Welk jaar is het? | 0,57 (0,50) | 0,52 | 0,60 (0,49) | 0,57 |
| 5 | Hoe heet de plaats waar u woont? | 0,92 (0,27) | 0,53 | 0,93 (0,26) | 0,62 |
| 8 | In welk jaar bent u geboren? | 0,97 (0,16) | 0,59 | 0,92 (0,27) | 0,58 |
| 9 | In welke maand bent u geboren? | 0,98 (0,13) | 0,84 | 0,98 (0,15) | 0,94 |
| 10 | Op de hoeveelste van de maand bent u geboren? | 0,97 (0,16) | 0,78 | 0,97 (0,16) | 0,82 |
| 11 | Wie is op het ogenblik onze koningin? | 0,78 (0,41) | 0,72 | 0,83 (0,38) | 0,65 |
| 12 | Wie was koningin vóór haar? | 0,69 (0,46) | 0,67 | 0,74 (0,44) | 0,67 |
| 13 | (Volgorde juist?) | 0,71 (0,45) | 0,64 | 0,76 (0,43) | 0,66 |
| 15 | Wie is op het ogenblik onze minister-president? | 0,46 (0,50) | 0,65 | 0,44 (0,50) | 0,61 |
| 16 | Wie was minister-president vóór hem? | 0,17 (0,37) | 0,82 | 0,21 (0,41) | 0,69 |
| 17 | (Volgorde juist?) | 0,15 (0,36) | 0,92 | 0,16 (0,37) | 0,81 |
| | M (SD) | 7,38 (2,31) | | 7,55 (2,33) | |
| | Mokkens H | | 0,67 | | 0,66 |
| | NIRT-model | | DMM | | DMM |
| | Rho | | 0,86 | | 0,86 |

NIRT-model = niet-parametrisch itemresponsetheoriemodel; DMM = dubbel monotoon model.

Welke itemverzameling voldoet aan dubbele monotonie?

Geen van de itemverzamelingen (twintig of veertien items) voldeed volledig aan de tweede monotonie-eis van het Mokkenmodel. In de 'training set' (de eerste willekeurig gekozen helft van de deelnemers) werden met behulp van het algoritme uit het MSP-programma elf items gevonden met een item-H-coëfficiënt en een schaalbaarheidscoëfficiënt $> 0,50$. Deze items zijn weergegeven in Tabel 3.

Vervolgens is met kruisvalidering de schaalbaarheid van deze items getoetst in een 'test set' (de andere helft van de steekproef). Tabel 3 laat zien dat de schaalbaarheidscoëfficiënten voor de elf items in de twee deelsteekproeven nauwelijks van elkaar verschillen ($H=0,67$, resp. $0,66$). De item-H-coëfficiënten zijn in de twee steekproeven niet altijd gelijk, maar de verschillen zijn klein, en lopen uiteen van $0,00$ tot $0,13$ per item.⁴ De items zijn deels (acht) afkomstig uit de CST14, deels (drie) uit de zes extra items van de CST20. De geselecteerde items betreffen het huidige jaar, de woonplaats, de geboortedatum van de onderzochte en de namen van vorstin en premier, inclusief die van hun ambtsvoorgangers. In beide steekproeven voldoet de itemverzameling

aan de tweede monotonie-eis: zij zijn, gegeven hun moeilijkheidsgraad, voor elke respondent op dezelfde wijze geordend.

Validiteit van de diverse itemcombinaties

Tabel 4 laat zien in hoeverre de verschillende verzamelingen van CST-items onderscheid mogelijk maken tussen deelnemers met een relatief hoog niveau van cognitief functioneren ($ADS_3 > 0$) en een minder hoog niveau ($ADS_3 \leq 0$).

De Pearson-correlatie tussen somscores op CST en ADS_3 geven een eerste indicatie van de samenhang. De correlaties variëren van $0,58$ (voor ADS_3 en CST-select) tot $0,67$ (voor ADS_3 en CST20). De correlaties verschillen echter niet significant van elkaar, gelet op de overlappende 95%-betrouwbaarheidsintervallen.¹³

De kruistabellen in Tabel 4 zijn het resultaat van discriminantanalyse. Voor elk van de drie CST-itemverzamelingen berekende de discriminantanalyse de CST-score waarmee deelnemers met een zo hoog mogelijke trefkans werden ondergebracht bij de groep met een ADS -score ≤ 0 ($n=255$), of de groep met een ADS -score > 0 ($n=201$). Het percentage correcte classificaties (CC) varieerde van $73,5\%$ (op basis van de CST20) tot $72,1\%$ (op basis van de CST14). Met de selectie

Tabel 4 Classificatie van deelnemers naar twee ADS_3 -niveaus door drie itemcombinaties van de Cognitieve Screening Test (CST).

| ADS ₃ | CST ₂₀ (0-20) | | | CST ₁₄ (0-14) | | | CST-Select (0-11) | | |
|------------------------------|--------------------------|------|-------------------|--------------------------|------|-------------------|-------------------|-----|-------------------|
| | ≤ 14 | > 14 | N | ≤ 11 | > 11 | N | ≤ 8 | > 8 | N |
| ≤ 0 | 176 | 79 | 255 | 171 | 84 | 255 | 215 | 40 | 255 |
| > 0 | 42 | 159 | 201 | 43 | 158 | 201 | 85 | 116 | 201 |
| N | 218 | 238 | 456 | 214 | 242 | 456 | 300 | 156 | 456 |
| Pearson-correlatie 95%-BI | | | 0,67 0,62-0,72 | | | 0,64 0,58-0,69 | | | 0,58 0,52-0,64 |
| CC | | | 73,5 | | | 72,1 | | | 72,6 |
| PU | | | 47,5 | | | 45,2 | | | 46,1 |
| PVW | | | 80,7 | | | 79,9 | | | 71,7 |
| NVW | | | 66,8 | | | 65,3 | | | 74,4 |
| Sensitiviteit | | | 69,0 | | | 67,1 | | | 84,3 |
| Specificiteit | | | 79,1 | | | 78,6 | | | 57,7 |

CC = correcte classificaties (%); PU = prognostic utility (%), berekend door het percentage van de gemiddelde voorspellende waarde $[(PVW+NVW)/2]$ te verminderen met 50 (de procentuele toevalskans bij een indeling in twee klassen), de uitkomst te delen door 50 (dat wil zeggen 100 minus de procentuele toevalskans), en deze uitkomst ten slotte te vermenigvuldigen met 100 om het percentage van de PU te verkrijgen.

PVW (%) = positief voorspellende waarde (van een lage CST-score voor een lage ADS -score); NVW (%) = negatief voorspellende waarde (van een hoge CST-score voor een hoge ADS -score); sensitiviteit (%) is het aantal deelnemers met een lage ADS -score dat ook een lage CST-score had; specificiteit (%) is het aantal deelnemers met een hoge ADS -score dat ook een hoge CST-score had.

van elf items werden 72,6% juiste classificaties bereikt. De *prognostic utility* (PU) komt met deze rangorde overeen. De PU voor toewijzing aan een van de twee ADS-niveaus was 47,5% voor de CST20. Dat wil zeggen: toepassing van de optimale CST20-grensscore gaf een 47,5% betere classificatie dan op grond van toeval zou zijn bereikt.

Zoals Tabel 4 laat zien, ging een lage score (≤ 14) op de CST20 in 81% van de gevallen gepaard met een lage score op de ADS3 (176/218). In de tabel wordt dit in navolging van de medische beslis-kunde de positief voorspellende waarde (PVW) genoemd, ofwel de kans dat iemand met een lage CST-score ook een lage score op de ADS heeft.¹⁴ De ‘voorspelling’ (op basis van een CST20-score > 14) van een hoge score op de ADS3 had een wat lagere trefkans van 67% (159/238). Dit wordt in Tabel 4 weergegeven als de negatief voorspel-lende waarde (NVW), ofwel de waarschijnlijkheid waarmee iemand met een hoge score op de CST ook een hoge score op de ADS3 had. In de tabel zijn voor de verschillende itemcombinaties van de CST ook de sensitiviteit en de specificiteit berekend. Sensitiviteit is het percentage deelnemers met een lage score op de ADS3 dat op een itemverzameling van de CST onder de grens-score kwam (bijvoorbeeld voor de CST20: 176/255=69,0%). Specificiteit is het percentage deelnemers met een hoge score op de ADS3 dat boven een CST-grensscore kwam (bijvoorbeeld voor de CST20: 159/201=79,1%).

De besliskundige parameters zijn voor de CST20 en de CST14 bijna gelijk. Gebeurde de verdeling over twee niveaus van de ADS3 op basis van de kleinste CST-itemverzameling met een optimale schaalbaarheid in termen van H, dan was de *prognostic utility* iets beter dan voor de standaard korte versie van de CST (CST14). De relatief hoge sensitiviteit (84%) ging echter gepaard met een verlaagde specificiteit (58%).

Discussie

In dit onderzoek is de itemreeks van de CST getoetst aan een non-parametrisch model uit de itemresponsetheorie (IRT). Tevens werd gezocht naar een itemreeks met een optimale schaalbaarheid, volgens een dubbel monotoon model. De analyses werden uitgevoerd op andere gegevens dan waarmee de CST oorspronkelijk werd ontwikkeld.¹⁵

De twintig items van de CST vormden, evenals de veertien items van de standaard korte versie, volgens de criteria van het schaalmodel van Mokken een niet al te sterke schaal. Met behulp van een geautomatiseerde zoekprocedure kon een verzameling van elf items worden gevonden die wel voldeed aan de criteria voor een sterke schaal. Vragen naar het huidige jaar, de woonplaats, de geboortedatum, de namen van de huidige en de voorgaande koningin, en die van de premiers vormen een eendimensionele, homogene schaal, waarvan de somscore op een be-

trouwbaar wijze individuele verschillen zichtbaar maakt in het niveau van cognitief functioneren. Bovendien is de somscore betrouwbaar te herleiden tot de invloed van een enkele latente eigenschap. Dat zijn voor de klinische praktijk waardevolle eigenschappen.^{16,17} Vanuit meet-theoretisch oogpunt heeft de clinicus met de verzameling van elf items een betrouwbaarder instrument om de positie van een respondent op de onderzochte latente eigenschap vast te stellen dan met andere itemverzamelingen die minder goed voldoen aan de voorwaarden van het schaalmodel van Mokken. In de voorspelling van een onafhankelijk criterium presteerde de verzameling van elf items, gelet op het aantal juiste classificaties en de *prognostic utility*, ongeveer even goed als de standaardversies van de CST.

Beperkingen

Dit onderzoek geeft geen antwoord op de vraag hoe de verzameling van geselecteerde items onderscheid mogelijk maakt tussen de diagnostische categorieën ‘dementie’ en ‘normale cognitieve veroudering’. Er waren immers geen deelnemers met een normale cognitie. In eerder onderzoek zijn de optimale grensscores voor het onderscheid tussen dementie en normale cognitieve veroudering vastgesteld op 10,1 en 12,1, voor respectievelijk de CST14 en de CST20.^{1,18-20}

Tegen de keuze van de ADS3 als onafhankelijk extern criterium kan worden ingebracht dat ADS en CST op verschillende dagen werden afgenomen. Dit leidde tot een klein verlies aan respondenten (2%), die op de dag dat afname van de ADS was gepland niet meer in dagbehandeling waren. Door afname op verschillende dagen heeft de dagdagelijkse variatie in cognitieve prestaties de correlatie tussen CST en ADS mogelijk sterker beïnvloed dan wanneer de vergelijking betrekking zou hebben gehad op tests die in dezelfde sessie waren afgenomen. Daarentegen werd de relatieve onafhankelijkheid van test (CST) en criterium (ADS) weer bevorderd door deze op afzonderlijke dagen af te nemen.

De hier gebruikte criteriumscore (ADS3) bevatte naast twee ongelijksoortige tests (*Visueel Geheugen* en *Fluency*) het onderdeel *Oriëntatie*. Twee van de drie vragen uit *Oriëntatie* (Welke maand is het? Welk jaar?) hebben betrekking op identieke items uit de CST. De scoringsprocedure van de ADS is echter iets soepeler dan die van de CST. Volgens de handleiding bij de ADS wordt vermelding van de voorafgaande maand goed gerekend, indien niet langer dan twee weken geleden, evenals vermelding van het voorafgaande jaar, indien niet langer dan twee maanden geleden. Het is niet uit te sluiten dat een deel van de gemeenschappelijke variantie in CST en ADS door enkele overlappende items wordt verklaard, maar gelet op het grote aantal CST-items en het aandeel van andere onderdelen van de ADS in de criteriumscore mag deze overlap aanvaardbaar klein worden geacht.

Literatuur

- 1 De Graaf A, Deelman BG. *Cognitieve Screening Test*. Lisse: Swets en Zeitlinger, 1991.
- 2 Teresi JA. Mini-Mental State Examination (MMSE): scaling the MMSE using item response theory (IRT). *Journal of Clinical Epidemiology* 2007;60:256-259.
- 3 Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use*. Fourth edition. Oxford: Oxford University Press, 2008.
- 4 Sijtsma K, Molenaar IW. *Introduction to nonparametric item response theory*. London: Sage, 2002.
- 5 Swanborn PG. *Schaaltechnieken. Theorie en praktijk van acht eenvoudige procedures*. Tweede druk. Meppel: Boom, 1988.
- 6 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, DC: American Psychiatric Association, 1994.
- 7 Lindeboom J, Jonker C. Een korte test voor dementie-screening. *Tijdschrift voor Gerontologie en Geriatrie* 1988;19:97-102.
- 8 Molenaar IW, Van Schuur WH, Sijtsma K, Mokken RJ. MSPWIN 5.0 A program for Mokken scale analysis for polytomous items. Groningen: Science Plus Group, 2002.
- 9 Morales LS, Flowers C, Gutierrez P, Kleinman M, Teresi JA. Item and scale differential functioning of the Mini-Mental State Exam assessed using the differential item and test functioning (DFIT) framework. *Med Care* 2006;44:S143-S151.
- 10 Van der Heijden PGM, Van Buuren S, Fekkes M, Radder J, Verrips E. Unidimensionality and reliability under Mokken scaling of the Dutch language version of the SF-36. *Quality of Life Research* 2003;12:189-198.
- 11 Lindeboom J, Jonker C. *Amsterdamse Dementie-Screeningstest*. Lisse: Swets and Zeitlinger, 1989.
- 12 Yarnold PR, Soltysik RC. *Optimal data analysis. A guidebook with software for Windows*. Washington DC: American Psychological Association, 2005.
- 13 Gardner MJ, Altman DG. *Statistics with confidence. Confidence intervals and statistical guidelines*. London: British Medical Journal, 1989.
- 14 De Koning I, Luteijn F. Psychometrie. In: Hendriks M, Kessels R, Gorissen M, Schmand B, editors. *Neuropsychologische diagnostiek*. Amsterdam: Boom, 2006: 103-118.
- 15 Maring W, Deelman BG. De Cognitieve Screening Test: kort en lang. *Tijdschrift voor Gerontologie en Geriatrie* 1999;30:205-211.
- 16 Hengeveld MW, Schudel WJ. *Het psychiatrisch onderzoek*. Utrecht: De Tijdstroom, 2003.
- 17 Kwaliteitsinstituut voor de Gezondheidszorg CBO. *Diagnostiek en medicamenteuze behandeling van dementie*. Utrecht: Kwaliteitsinstituut voor de Gezondheidszorg CBO, 2005.
- 18 Ponds RWHM, Verhey FRJ, Rozendaal N, Jolles J, Deelman BG. Dementie-screening; validiteit van de Cognitieve Screeningstest (CST) en de Mini-Mental State Examination (MMSE). *Tijdschrift voor Gerontologie en Geriatrie* 1992;23:94-99.
- 19 Schmand B, Deelman BG, Hooijer C, Jonker C, Lindeboom J. De item-reeks van de Cognitieve Screening Test vergeleken met die van de Mini-Mental State Examination. *Tijdschrift voor Gerontologie en Geriatrie* 1996;27:29-33.
- 20 Elzen H, Schmidt I, Bouma A. De diagnostische waarde van de kloktekening in de geriatrie. *Tijdschrift voor Gerontologie en Geriatrie* 2004;35:107-113.
- 21 Verhage F. *Intelligentie en leeftijd bij volwassenen en bejaarden*. Assen: Van Gorcum, 1964.