

# Letterfluency: psychometrische eigenschappen en Nederlandse normen

B. Schmand<sup>a</sup>, S.C. Groenink<sup>b</sup>, M. van den Dungen<sup>c</sup>

## Letterfluency: psychometric properties and Dutch normative data

Normative data were collected for a Dutch version of the Controlled Oral Word Association Test (COWAT) in 200 healthy subjects between 17 and 89 years of age. The COWAT is a letterfluency task that is widely used in clinical neuropsychology. Fluency is an important aspect of executive functioning. The psychometric properties of the Dutch version of the test were largely comparable to those of the original COWAT. Its reliability is 0,80, and its scores are significantly related to level of education and/or vocabulary, but not to age or gender. A regression formula is provided by which the raw scores can be corrected for level of education.

Keywords: verbal fluency, norms, neuropsychological tests.

*Tijdschr Gerontol Geriat 2008; 39: 64-76*

---

## Samenvatting

Een Nederlandstalige bewerking van de Controlled Oral Word Association Test (COWAT) werd genormeerd bij 200 gezonde personen in de leeftijd van 17 tot 89 jaar. De COWAT is een letterfluency of fonemische fluencytest,

<sup>a</sup> Universiteit van Amsterdam, programmagroep Psychonomie en Academisch Medisch Centrum, afdeling Neurologie

<sup>b</sup> Universiteit van Tilburg, Faculteit Sociale Wetenschappen

<sup>c</sup> Psycholoog, Sophia Revalidatie, Den Haag  
Correspondentie: B. Schmand AMC Neurologie H2 Postbus 22660 1100DD Amsterdam. T: 020-566 3590  
E: b.schmand@amc.uva.nl

die wereldwijd veel wordt gebruikt in de neuropsychologie. Fluency is een belangrijk aspect van het executief functioneren.

De psychometrische eigenschappen (normen, betrouwbaarheid, validiteit) van de Nederlandse versie kwamen goed overeen met die van de oorspronkelijke test. De test heeft een betrouwbaarheid van 0,80 (interne consistentie; paralleltest-betrouwbaarheid is 0,78), en de score hangt significant samen met het opleidingsniveau en/of de woordkennis, maar niet met de leeftijd en het geslacht van de onderzochte persoon.

In een appendix wordt een regressieformule gegeven waarmee de ruwe score kan worden gecorrigeerd voor opleiding.

---

## Inleiding

Tests van de verbale fluency of woordvlotheid worden in vrijwel elk neuropsychologisch onderzoek gebruikt, of het nu gaat om onderzoek van oudere patiënten in een geheugenpoli, kinderen met leerstoornissen, of jongvolwassen patiënten met traumatisch hersenletsel.

Bij een verbale fluencytest moet de patiënt in een beperkte tijd zoveel mogelijk woorden opnoemen die aan een bepaald criterium voldoen, bijvoorbeeld zoveel mogelijk dieren in één minuut. Fluencytests zijn er in een groot aantal varianten, zowel verbale als nonverbale.

Thurstone was een van de eersten die een verbale fluencytest toepaste. Hij zag 'word fluency' als één van de 'primary mental abilities'.<sup>1</sup> Proefpersonen moesten bij deze test zoveel mogelijk woorden opschrijven uit een bepaalde categorie. Deze opdracht staat tegenwoordig bekend als categoriefluency. Guilford plaatste verbale fluency in het bredere kader van "divergent denken", dat wil zeggen denken dat in verschillende richtingen verloopt en meerdere, of originele oplossingen kan opleveren.<sup>2</sup> Divergent denken impliceert wendbaarheid van geest en is dus een belangrijk bestanddeel van creativiteit. Guilford's fluencytests hadden dan ook niet alleen betrekking op het genereren van woorden. Hij gaf zijn proefpersonen allerlei divergente denkopdrachten. Hij liet hen bijvoorbeeld zoveel mogelijk nieuwe toepassingen bedenken van alledaagse voorwerpen (Wat kun je nog meer doen met een balpen dan schrijven of tekenen?), of zoveel mogelijk gevolgen van een onwaarschijnlijke premisse (Wat zouden de gevolgen zijn als de mensen niet meer hoefden te eten?).

In de hedendaagse neuropsychologie worden drie soorten fluencytests veel gebruikt, namelijk categoriefluency, letterfluency en figuurfluency. Bij categoriefluency is het de bedoeling dat de proefpersoon zoveel mogelijk woorden uit een bepaalde categorie noemt (of opschrijft). Veel gebruikte categorieën zijn dieren, beroepen en dingen die je in de supermarkt kunt kopen. Er bestaan lijsten van tientallen categorieën met tellingen van de woorden die mensen produceren bij elke categorie.<sup>3-5</sup> De fluencytest waarbij het criterium is dat de te produceren woorden met een bepaalde letter moeten beginnen wordt fonemische, fonologische of letterfluency genoemd. De derde soort fluencytest, figuurfluency, is een nonverbale test, waarbij zoveel mogelijk verschillende figuurtjes moeten worden getekend.<sup>6-8</sup>

Fluencytests zijn populair omdat ze snel zijn af te nemen, betrouwbare scores opleveren, en sensitief zijn voor allerlei cerebrale functiestoornissen. Bovendien zijn vooral de verbale fluencytaken op ontelbare manieren gevalideerd in de differentiële psychologie, de klinische neuropsychologie, en in andere neurowetenschappen.<sup>9-19</sup> In de differentiële psychologie worden fluencytaken gebruikt als on-

derdeel van intelligentietests en creativiteitstests. In de neurowetenschappen worden fluencytaken vooral gezien als een operationalisatie van de executieve functies. Immers, bij een fluencyopdracht moet de proefpersoon nieuw gedrag initiëren, de spelregels van de opdracht respecteren, een verstandige zoekstrategie ontwerpen en deze strategie volgen, overschakelen naar een andere strategie als de huidige te weinig oplevert, inadequate responsen inhiberen, en controleren of zijn responsen aan het criterium voldoen.<sup>14, 20</sup> Laesiestudies hebben laten zien dat patiënten met frontale hersenbeschadigingen slecht presteren op fluencytaken.<sup>14, 17</sup> Onderzoek met functionele beeldvorming bij gezonde personen laat onveranderlijk zien dat fluencytaken de prefrontale cortex activeren, en soms komt ook naar voren dat de gyrus cinguli anterior actief is.<sup>9, 21</sup>

Ook het werkgeheugen is belangrijk bij fluencytaken, want men moet de opdracht (criteria waaraan reponsen moeten voldoen) paraat houden, bijhouden waar men al heeft gezocht, en bijhouden welke responsen men al gegeven heeft. Overigens wordt het werkgeheugen vaak gezien als deel van de uitvoerende functies (zie bijvoorbeeld Gazzaniga et al. 2002, pg 514-524).<sup>22</sup> Bij verbale fluencytests spelen bovendien talige aspecten een belangrijke rol. Verbale fluencytests zijn daarom niet alleen tests van de executieve functies, maar ook van het semantische geheugen (vooral categoriefluency) en van het lexicale geheugen (vooral letterfluency). Ze zijn daardoor geschikt om fatische stoornissen op het spoor te komen, zoals problemen met woordvinding en degeneratie van semantische kennis. Vooral patiënten met schade in de linker temporaalkwab en patiënten met de ziekte van Alzheimer presteren slecht op categoriefluency.<sup>16, 20</sup> Ten slotte is het snelheidskarakter een belangrijk aspect van fluencytaken. Aan de proefpersoon wordt immers gevraagd zo snel mogelijk te werken. De score reflecteert deels dus ook de snelheid waarmee items worden gezocht in en opgediept uit het mentale lexicon, en de snelheid van responderen, dat wil zeggen de snelheid van articuleren, schrijven of tekenen.

Voor het Nederlandse taalgebied bestaat een goed genormeerde categoriefluencytest, namelijk de subtest Woordopnoemen van de Groninger Intelligentie Test.<sup>23</sup> De normen van de GIT zijn gestratificeerd naar leeftijd. Van deze test zijn bovendien normgegevens uit de Maastricht Aging Study gepubliceerd, die behalve een correctie voor leeftijd ook een correctie voor opleidingsniveau en geslacht mogelijk maken.<sup>24</sup> De figuurfluencytest van Ruff is ook recent genormeerd voor Nederland.<sup>6</sup> In Vlaanderen is een letterfluencytest beschikbaar,<sup>25</sup> maar een goed genormeerde letterfluencytest voor Nederland bestaat nog niet. Weliswaar hebben Van der Elst et al. normen gepresenteerd voor een letterfluencytaak, maar hun versie van de test bestond eruit dat men in één minuut zoveel mogelijk woorden

van vier letters beginnend met een M moest opnoemen.<sup>24</sup> Deze test heeft echter twee bezwaren. Ten eerste is hij zo kort dat niet verwacht mag worden dat hij erg betrouwbaar is. Van der Elst et al. melden geen gegevens over de betrouwbaarheid. Ten tweede is het formaat van de test niet gangbaar ook al is hij ontleend aan het werk van Thurstone.<sup>1</sup> De variant die internationaal het meest gebruikt wordt, is de Controlled Oral Word Association Test (COWAT) uit de Multilingual Aphasia Examination.<sup>26</sup> Bij deze test worden drie trials van één minuut afgenomen waarbij woorden moeten worden opgenoemd die met een bepaalde letter beginnen. Deze variant staat ook bekend als de FAS test; dit waren de drie letters die oorspronkelijk in de COWAT werden gebruikt.

Wij hebben een Nederlandse letterfluencytest geconstrueerd, die overeenkomt met de COWAT en dus aansluit bij wat internationaal gangbaar is. In deze bijdrage beschrijven we de constructie en de psychometrische eigenschappen van de test (normen, betrouwbaarheid en hertest-effect, en validiteit). Wat betreft de validiteit, dat wil zeggen de vraag welke vaardigheden de test precies meet, proberen we te onderzoeken welke cognitieve vaardigheden bijdragen aan de uitvoering van de letterfluencytaak. We betrekken daarbij andere soorten fluency, evenals de woordenschat en enkele aspecten van de uitvoerende functies. Ook onderzoeken we de veronderstelling dat de productie aan het begin van de taak vooral een beroep doet op automatische processen van informatieverwerking, terwijl daarna de woordproductie tot stand komt door meer gecontroleerde zoekprocessen in het lexicale en semantische geheugen.<sup>27, 28</sup>

## Methoden

### *Constructie [1]*

In deze context hangt het aantal woorden dat kan worden geproduceerd af van de frequentie waarmee de betreffende letter als beginletter van woorden voorkomt in de taal. De keuze van de letters is daarom gebaseerd op een schatting van de aantallen woorden met dezelfde beginletter in de officiële woordenlijst van de Nederlandse taal (het groene boekje) en in het Winkler Prins woordenboek. Een vergelijkbare methode werd gevolgd door Benton en Hamsher bij de keuze van de COWAT letters.<sup>26</sup> Voor elke letter (behalve de Q, de X en de Y) werd het aantal bladzijden geteld dat de betreffende letter beslaat. Deze tellingen op grond van beide boeken correleren zeer hoog ( $r=0,92$ ). Hetzelfde werd gedaan voor

Van Dale's Groot Woordenboek van de Nederlandse taal, maar die telling correleerde veel lager met de andere twee (0,55 en 0,67), en werd daarom niet gebruikt. Vervolgens werd een ordening van letter-moeilijkheid gemaakt door de tellingen op grond van het groene boekje en het woordenboek te z-transformeren en te sommeren. De letter S bleek veruit de gemakkelijkste letter te zijn. De volgorde van de volgende negen letters was P-A-K-O-G-T-D-M-R. Met deze negen letters werden drie versies van de test gemaakt. De uiteindelijke keuze van de te gebruiken lettercombinaties werd gebaseerd op deze volgorde in moeilijkheid: DAT, KOM en PGR. De vergelijkbaarheid van deze drie letter-combinaties werd getoetst in een groep HBO-studenten en artsen die de test groepsgewijs schriftelijk aflegden ( $n=40$ ). De versies bleken vergelijkbaar te zijn wat betreft het gemiddelde aantal geproduceerde woorden. (Deze gegevens werden niet gebruikt bij het samenstellen van de normen.) Vervolgens werd de COWAT instructie vertaald (zie Appendix A).

### *Proefpersonen*

In het najaar van 2003 werden normgegevens verzameld bij 200 gezonde vrijwilligers uit de algemene bevolking in Amsterdam en omgeving. Zij werden gerekruteerd door middel van de sneeuwbal methode: de onderzoekers begonnen met het testen van familieleden en vrienden, en vroegen aan elke onderzochte of hij nog meer gegadigden kende, die vervolgens werden uitgenodigd voor deelname. Proefpersonen die het Nederlands niet als moedertaal hadden, en proefpersonen met psychiatrische aandoeningen of ziekten van het centrale zenuwstelsel werden uitgesloten.

De proefpersonen varieerden in leeftijd tussen 17 en 89 jaar (gemiddeld 53,2 SD=17,9).

Van hen waren 60,5% vrouwen. Het hoogst behaalde opleidingsniveau varieerde van onvolledig basisonderwijs tot afgerond universitair onderwijs (tabel 1). Het mediane onderwijsniveau was mulo/mavo/mbo.

De groep was niet geheel representatief voor de bevolking vergeleken met CBS gegevens over 2001.<sup>29</sup> Mensen met havo en hogere opleiding waren oververtegenwoordigd (ongeveer 45% i.p.v. 30% in de leeftijdsrange tot 65 jaar). Ook de ouderen waren te sterk vertegenwoordigd (22% i.p.v. 9% in de leeftijd van 70 jaar en ouder). Voor deze oververtegenwoordiging van ouderen werd bewust gekozen omdat naar verwachting de test in de kliniek vooral bij ouderen zal worden toegepast.

Tabel 1 Verdeling van de proefpersonen over onderwijsniveaus en leeftijdscategorieën.								
	< 6 jaar lagere school	6 klassen lagere school	> 6 jaar lagere school	lbo	mulo, mavo, mbo	havo, vwo, hbo	universiteit	totaal
17-19 jaar	0	0	0	1	4	1	0	6
20-29 jaar	0	0	0	1	10	5	5	21
30-39 jaar	0	0	0	0	5	11	4	20
40-49 jaar	0	0	0	6	11	13	2	32
50-59 jaar	0	0	2	6	16	21	4	49
60-69 jaar	0	1	1	9	8	8	1	28
70-79 jaar	1	5	2	5	9	6	3	31
80-89 jaar	1	0	2	4	4	1	1	13
totaal	2	6	7	32	67	66	20	200

lbo= lager beroepsonderwijs, mulo= meer uitgebreid lager onderwijs, mavo= middelbaar algemeen voortgezet onderwijs, mbo= middelbaar beroepsonderwijs, havo= hoger algemeen voortgezet onderwijs, vwo= voortgezet wetenschappelijk onderwijs, hbo= hoger beroepsonderwijs

## Tests

**Letterfluency.** De test wordt mondeling afgenomen. De proefpersoon moet drie maal gedurende één minuut zoveel mogelijk woorden opnoemen die beginnen met een bepaalde letter, te weten de letters van één van de bovengenoemde lettercombinaties. In de instructie worden nog enkele beperkende regels opgelegd die daarbij gevolgd moeten worden (zie Appendix A). Per letter werd het aantal correct opgenoemde woorden geteld. De letterfluency score was het totale aantal correct genoemde woorden gesommeerd over de drie letters. Woorden die meer dan één keer werden genoemd, woorden die niet aan de regels van de testinstructie voldeden (zie Appendix A), en niet-bestaande woorden werden fout gerekend.

Behalve de letterfluencytaak werden de volgende tests afgenomen.

**Categoriefluency.**<sup>23</sup> De proefpersoon werd gevraagd in één minuut zoveel mogelijk verschillende dieren op te noemen. Dit werd herhaald met de categorieën beroepen, en artikelen die in een supermarkt te koop zijn.

**Ruff Figural Fluency Test (FFT).**<sup>8</sup> Bij deze test moet de proefpersoon in vijf minuten zoveel mogelijk verschillende figuurtjes tekenen, waarbij hij zich aan een aantal regels moet houden.

**Boston Naming Test (BNT).**<sup>30-31</sup> Bij deze test moeten 60 lijntekeningen van voorwerpen en dieren worden benoemd.

**Stroop kleur-woord test.**<sup>32</sup> Bij deze test krijgt de proefpersoon drie kaarten voorgelegd met elk 100 stimuli. Op de eerste kaart staan namen van

kleuren die in zwarte inkt zijn gedrukt. De opdracht is de kleurnamen zo snel mogelijk op te lezen. Op de tweede kaart staan dezelfde kleuren, maar dan in gekleurde vlakjes. De opdracht is de kleuren zo snel mogelijk op te noemen. Op de derde kaart staan weer namen van kleuren, maar deze keer in gekleurde inkt gedrukt. De kleur van de inkt komt niet overeen met de naam van de kleur, bijvoorbeeld het woord 'rood' gedrukt in groene inkt. De opdracht is de kleuren van de inkt te noemen en de kleurwoorden te negeren. Dit is een test van de selectieve aandacht en van het vermogen tot inhibitie van voor de hand liggende maar niet adequate responsen. De score is de tijd in seconden die men nodig heeft om de kleuren op te lezen c.q. te benoemen. Deze ruwe scores kunnen worden omgezet in T-scores die gecorrigeerd zijn voor leeftijd, opleiding en geslacht.<sup>33</sup>

**Nederlandse Leestest voor Volwassenen (NLV).**<sup>34</sup> Bij deze test moeten onregelmatig gespelde woorden worden opgelezen. Het aantal correct uitgesproken woorden geeft een schatting van de uitgebreidheid van de woordenschat en de verbale intelligentie.

**Seven minute screen.**<sup>35,36,53</sup> Dit is een korte dementie-screeningstest die alleen bij de deelnemers van 60 jaar en ouder werd afgenomen om dementie uit te sluiten.

## Procedure

De drie versies van de letterfluency test (DAT, KOM, PGR) werden in gebalanceerde volgorde afgenomen bij 152 personen. Bij 48 personen

werd alleen de versie DAT afgenomen. Binnen elke versie werden de drie letters steeds in dezelfde volgorde aangeboden. Tussen elke versie werd een interval aangehouden van steeds 5 à 10 minuten, waarin een andere test werd afgenomen (Stroop kleur-woord test en Ruff FFT). Bij de helft van de groep ( $n=100$ ) werden tevens de andere genoemde tests afgenomen. Voor deze personen duurde het onderzoek ongeveer een uur. Bij hen werd aan het eind van de sessie de eerst afgenomen versie van de letterfluencytest opnieuw afgenomen.

#### *Statistische analyses*

Allereerst werd nagegaan of de drie versies parallel zijn. Daartoe werd onderzocht of ze even moeilijk zijn door de gemiddelde scores van de drie versies bij eerste afname te vergelijken met behulp van covariantieanalyse, met leeftijd en opleidingsniveau als covariaten en geslacht als extra factor. De vorm van de scoreverdelingen werd onderzocht door de skewness en kurtosis te berekenen en te toetsen of deze afwijken van de normale verdeling door middel van de Kolmogorov-Smirnov toets. Vervolgens werd nagegaan of de betrouwbaarheid van de drie versies vergelijkbaar is door de interne consistentie en de paralleltestbetrouwbaarheid te berekenen. Bij gebleken paralleliteit werden de overige analyses gedaan zonder rekening te houden met de versie.

Hertest-effecten werden onderzocht door de toename van de scores bij herhaalde afname van de test te berekenen en de verschillen te toetsen met gepaarde t-toetsen.

Voor de presentatie van de normgegevens werd een regressie-benadering gevolgd: met behulp van lineaire regressie werd een formule opgesteld waarmee berekend kan worden wat de verwachte score is op basis van relevante kenmerken van de persoon. Het verschil tussen de verwachte score en de feitelijk behaalde score werd uitgedrukt in een standardscore of in een percentielscore die gecorrigeerd is voor de betreffende kenmerken. Deze worden weergegeven in tabelvorm.

De validiteitsvraag werd onderzocht door correlaties met andere fluencymaten, de NLV, de Boston naming test, en de Stroop test te berekenen. Om na te gaan of de taak in het begin andere vaardigheden aanspreekt dan later in de minuut werden deze correlaties ook voor de eerste 15 seconden van elke minuut en de rest van elke minuut apart berekend.

## **Resultaten**

### *Moeilijkheidsgraad en betrouwbaarheid*

Er werden gemiddeld 38,1 (SD=10,5; range 13 - 69) woorden opgenoemd bij de eerste afname

van de letterfluencytest. De drie versies van de test verschilden niet wat betreft de gemiddelde aantallen geproduceerde woorden (DAT: 37,0 SD=10,0; KOM: 38,3 SD=10,3; PGR: 39,9 SD=11,5;  $F(2,192)=1,54$ ,  $p=0,22$ ). Het opleidingsniveau had een significante invloed op de scores ( $F(1,192)=18,68$ ,  $p<0,001$ ), maar leeftijd en geslacht niet ( $p=0,86$  respectievelijk  $p=0,31$ ). De scores volgden de normale verdeling (skewness tussen 0,05 en 0,40; kurtosis tussen 0,01 en 0,34; Kolmogorov-Smirnov toets:  $p>0,38$ ).

De intercorrelaties van de scores bij elk van de negen letters lagen tussen 0,43 en 0,73; de gemiddelde correlatie was 0,58. De interne consistentie (Cronbachs alfa) werd bepaald door de drie letters van elke versie als aparte items te beschouwen. De interne consistentie lag tussen 0,80 (voor de versie KOM;  $n=152$ ) en 0,84 (voor de versie PGR;  $n=152$ ). De interne consistentie van de versie DAT was 0,82 ( $n=200$ ). De correlaties tussen de versies lagen tussen 0,78 en 0,81; dit is op te vatten als de paralleltest betrouwbaarheid.

Zowel de moeilijkheidsgraad als de betrouwbaarheid waren dus ongeveer hetzelfde voor de drie versies van de test. Daarom kunnen de versies als parallel worden beschouwd.

### *Hertesteffecten*

De 152 personen die alle versies van de test deden, behaalden de tweede keer een score die gemiddeld 3,9 woorden hoger lag dan bij de eerste afname (tweede afname 41,9 SD=11,1  $p<0,001$ ). De derde afname leverde gemiddeld nog 0,8 woorden meer op, maar het verschil met de tweede afname was niet significant (derde afname 42,7 SD=10,9  $p=0,16$ ). Voor de praktijk betekent dit dat het hertest effect vier woorden bedraagt, in ieder geval bij twee metingen zeer kort na elkaar met verschillende versies. Degenen ( $n=100$ ) die aan het eind van het onderzoek na ongeveer een uur nog een vierde afname deden met de eerst aangeboden versie, behaalden 1,9 woorden meer dan bij de derde afname ( $p=0,004$ ). Bij hen was de totale toename van de score over de vier afnamen 7,2 woorden (tabel 2). Een covariantie-analyse met opleiding als covariaat, de afgenomen versie als between subjects factor en de herhaling als within subjects factor liet wederom significante effecten van opleiding ( $F(1,96)=9,40$ ,  $p<0,003$ ) en van herhaling zien ( $F(1,96)=7,27$ ,  $p<0,008$ ), maar niet van de aangeboden versie ( $F<1$ ).

Tabel 2		Gemiddelde scores (standaarddeviatie) bij de eerste en de vierde afname van de letterfluencytaak.			
	DAT (N = 39)	KOM (N= 33)	PGR (N= 28)	totaal (N = 100)	
1 <sup>e</sup> afname	37,5 (10,2)	39,4 (9,8)	41,3 (13,3)	39,2 (11,0)	
4 <sup>e</sup> afname	45,5 (10,3)	46,5 (10,1)	47,6 (13,2)	46,4 (11,1)	

### Normen

Leeftijd en geslacht waren bij de eerste afname van de test niet van invloed op de score, zoals boven beschreven. Wel correleerde de score met het opleidingsniveau (Spearman's rho = 0,32;  $p < 0,001$ ). Bij de interpretatie van een individuele score dient men dus rekening te houden met het opleidingsniveau van de persoon in kwestie. Met een lineaire regressieanalyse werd een formule opgesteld waarmee de behaalde score van een individuele persoon kan worden omgezet in een normscore die gecorrigeerd is voor het opleidingsniveau. Op basis van deze formule werd tabel 3 opgesteld.

Slechts een minderheid van de proefpersonen (29%) deed de taak bij de eerste afname foutloos, en 32% maakte één fout. Gemiddeld werden er 0,9 (SD=1,1) woorden dubbel genoemd, en waren 0,8 (SD=1,1) woorden overtredingen van één van de andere regels. Meer dan drie keer een woord

dubbel noemen, of meer dan drie andere regel-overtredingen was echter uitzonderlijk (kwam bij 3,3% respectievelijk 2,7% van de proefpersonen voor). Het aantal dubbelen en andere regel-overtredingen nam niet significant toe of af bij de volgende afnamen en het correleerde niet met leeftijd of opleiding.

### Validiteit

De correlatie tussen letterfluency en categoriefluency was betrekkelijk hoog:  $r = 0,55$  met het aantal genoemde dieren,  $r = 0,54$  met het aantal beroepen en  $r = 0,42$  met het aantal supermarkt-artikelen ( $p < .001$ ; tabel 4). De correlatie met Ruff Figural Fluency Test was 0,30 ( $p = .003$ ). De verbanden met categoriefluency hielden stand na corrigeren voor de effecten van leeftijd en opleiding (partiële correlaties respectievelijk 0,50, 0,46 en 0,37  $p < .001$ ), maar het verband met de

Tabel 3		Omzettingstabel van ruwe letterfluency scores naar percentielscores en T-scores per opleidingsniveau.						
	< basis-school	basisschool	< lbo	lbo, vmbo-b	mbo, mulo, mavo, vmbo-t	havo, vwo, hbo	universiteit	
percentiel								T-score
99	50	53	55	58	61	63	65	73
95	44	47	49	52	55	57	59	67
90	40	43	45	48	50	53	55	63
75	34	37	39	42	44	47	49	57
50	27	30	32	35	37	40	42	50
25	20	23	25	28	30	33	35	43
10	14	17	19	22	24	27	29	37
5	10	13	15	18	20	23	25	33
1	4	7	9	12	14	17	19	27

T-scores zijn normaal verdeeld met een gemiddelde van 50 en een SD van 10. Voorbeeld: een patiënt met een lagere beroepsopleiding behaalt een score van 22 op de letterfluencytest. Zijn T-score is 37. Dit komt overeen met het 10<sup>e</sup> percentiel. Zie Appendix B voor details.



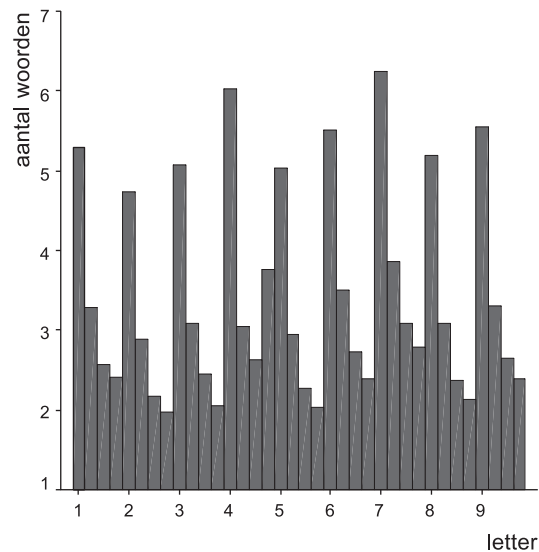
Ruff test werd aanzienlijk afgezwakt (partiële  $r = 0,19$   $p = .03$  eenzijdig).

Letterfluency correleerde ook met de ruwe scores op de Stroop kleur-woord test (tussen  $-0,30$  en  $-0,38$   $p < .002$ ). Het verband met de Stroop taak verdween echter grotendeels na correctie voor leeftijd, geslacht en opleiding (woordkaart:  $r = 0,25$   $p = .01$ ; kleurkaart:  $r = 0,15$  niet significant; kleur-woordkaart:  $r = 0,15$  niet significant).

Met de Boston Naming Test was er geen verband van betekenis. De correlatie met de NLV score was  $0,37$  ( $p < .001$ ). Deze correlatie was  $0,24$  ( $p = .02$ ) als werd gecorrigeerd voor het opleidingsniveau. Omgekeerd verdween de correlatie met opleiding als werd gecorrigeerd voor NLV score ( $r = 0,15$   $p = .14$ ).

De correlaties tussen de aantallen genoemde woorden in de eerste 15 seconden en in het resterende deel van de minuut enerzijds, en woordkennis en executief functioneren anderzijds werden berekend om na te gaan welke van deze factoren de productie in de verschillende fasen van de taak bepalen. De prestatie gedurende de eerste 15 seconden van de letters bij de eerste afname correleerde  $0,37$  ( $p < .001$ ) met de NLV en  $-0,42$  ( $p < .001$ ) met de Stroop kleur-woord kaart ( $-0,29$  [ $p < .004$ ] wanneer voor leeftijd en opleiding wordt gecorrigeerd). Voor de resterende 45 seconden van de eerste drie letters waren deze correlaties  $0,34$  ( $p < .001$ ) en  $-0,33$  ( $-0,23$  [ $p = .02$ ] na correctie voor leeftijd en opleiding). Deze correlaties verschilden niet significant van elkaar (Fis-

her toets,  $z \leq 1,49$   $p > .07$ ). Zowel de Stroop test als de NLV correleerde dus matig met de woordproductie gedurende de hele minuut.



Figuur 1. Gemiddeld aantal geproduceerde woorden. De nummers 1-9 op de horizontale as geven de presentatie van de letters aan. Per letter wordt de productie per periode van 15 seconden weergegeven.

Het gemiddelde verloop van de woordproductie is te zien in figuur 1. In de eerste 15 seconden van elke minuut werden 5 à 6 woorden genoemd, waarna de productie telkens aanzienlijk

**Tabel 4. Gemiddelde scores (standaarddeviaties) van de eerste afname van de letterfluencytest en de overige tests, en correlaties tussen letterfluency en de overige tests (N = 100).**

Test	gemiddeld (sd)	letterfluency	eerste 15 sec. <sup>a</sup>	rest min. <sup>b</sup>
Letterfluency	39,0 (11,1)	-	0,87**	0,96**
eerste 15 sec.	15,4 (4,2)	0,87**	-	0,71**
rest van minuut	23,6 (7,7)	0,96**	0,71**	-
aantal dieren	23,2 (5,4)	0,55**	0,56**	0,49**
aantal beroepen	17,4 (4,1)	0,54**	0,50**	0,50**
aantal supermarkartikelen	24,6 (5,7)	0,42**	0,45**	0,35**
Ruff figuurfluency	94,1 (22,1)	0,30*	0,42**	0,21*
NLV (ruwe score)	87,6 (8,8)	0,37**	0,37**	0,34**
Boston naming test	56,4 (2,6)	0,15	0,20	0,11
Stroop test woord sec.	39,6 (6,8)	-0,38**	-0,39**	-0,34**
Stroop test kleur sec.	53,6 (9,0)	-0,30*	-0,32**	-0,26*
Stroop test kleur-woord sec.	99,1 (27,3)	-0,39**	-0,42**	-0,33**

\* = significant op 0,05-niveau; \*\* = significant op 0,01-niveau (2-zijdig). a: Pearson-correlaties met de som van de woorden die in de eerste 15 seconden van elke minuut zijn genoemd. b: Pearson-correlaties met de som van de woorden die in de rest van elke minuut zijn genoemd.

afnam. Het hertesteffect is ook enigszins waarneembaar: de productie vanaf de vierde letter was iets hoger dan bij de eerste drie letters.

## Discussie

De psychometrische kenmerken van deze Nederlandse letterfluencytest komen in grote lijnen overeen met wat er over de Controlled Oral Word Association Test is gepubliceerd.<sup>17, 26, 37, 38</sup> Gemiddeld noemen volwassen personen ongeveer 13 woorden per minuut, waarvan de helft in de eerste 15 seconden. Daarna vlakke de prestatie af tot twee à drie woorden per 15 seconden. De meeste mensen maken één of twee fouten, ongeacht hun leeftijd en opleidingsniveau. De drie versies die we construeerden bleken parallel te zijn (vergelijkbaar wat betreft moeilijkheid en betrouwbaarheid), en zijn dus onderling uitwisselbaar.

De normen van de Nederlandse versie zijn dan ook vergelijkbaar met die van de COWAT.<sup>17,37</sup> Er is een verband tussen de prestatie en het opleidingsniveau ( $\rho=0,32$ ), al is het een bescheiden verband dat slechts een klein deel van de variantie in letterfluency scores verklaart. Dit verband lijkt te worden gemedieerd door de woordenschat. Immers, het verband met opleiding wordt verwaarloosbaar na correctie voor de uitgebreidheid van de woordenschat (NLV). Mannen en vrouwen verschillen niet wat betreft hun gemiddelde prestaties. Er is nauwelijks of geen relatie met de leeftijd, althans niet voor het door ons onderzochte leeftijdsbereik van volwassenen en ouderen, mits rekening wordt gehouden met het opleidingsniveau. Bij kinderen en adolescenten ligt de relatie met leeftijd uiteraard anders. Kinderen van negen jaar, bijvoorbeeld, noemen slechts ongeveer de helft van het aantal woorden dat volwassenen noemen.<sup>28</sup> De prestatie van kinderen stijgt zeer snel met het toenemen van de leeftijd.<sup>38</sup>

In tegenstelling tot een compilatie van normen uit een groot aantal andere studies<sup>39</sup> en een meta-analyse van studies waarin de COWAT is gebruikt<sup>40</sup> vonden wij geen effect van de leeftijd, althans niet in de range die wij onderzochten. Beide studies melden een afname van de prestatie met ruim een standaarddeviatie (10 à 15 woorden) over vijf decaden na de vroege volwassenheid. Deze discrepantie met onze gegevens is waarschijnlijk te wijten aan het feit dat in deze beide studies nauwelijks of geen rekening werd gehouden met het opleidingsniveau. Bovendien maakten beide onderzoeken gebruik van cross-sectionele gegevens. Dergelijke gegevens zijn onderhevig aan cohort effecten. Eén van die effecten is de verandering van het gemiddelde opleidingsniveau. Dat is in vrijwel de gehele wereld voor oudere cohorten lager dan voor jongere doordat het gemiddelde onderwijsniveau in de loop der jaren voortdurend is gestegen. Dus de afname van de prestatie op letterfluencytests

met het toenemen van de leeftijd is wellicht een schijnbare afname, die niet te wijten is aan veroudering op zich, maar aan het lagere opleidingsniveau van ouderen. Als wij geen rekening zouden hebben gehouden met opleiding, zouden we eenzelfde schijnbare afname met de leeftijd hebben gevonden. De gemiddelde prestatie is stabiel (rond de 40 woorden) tot ongeveer het 60e jaar, en daarna treedt een daling in tot ruim 30 woorden bij mensen van 80 jaar en ouder. Het gemiddelde opleidingsniveau volgt echter hetzelfde patroon (in ons onderzoek is de rangcorrelatie tussen leeftijd en opleiding  $-0,30$ ,  $p < 0,001$ ).

De betrouwbaarheid van de test is 0,80. Het hertest effect is ongeveer vier woorden, en ongeveer zeven woorden als men dezelfde versie twee keer gebruikt (zie Appendix C voor de praktische toepassing van deze gegevens). Uit de literatuur is bekend dat de lengte van het interval weinig of geen invloed heeft op de hertest betrouwbaarheid en het hertest effect.<sup>38</sup> We nemen aan dat dit voor deze Nederlandse versie eveneens het geval is, ook al hebben we het nog niet onderzocht bij gezonde vrijwilligers.

De betrouwbaarheid van letterfluency tests die uit slechts één trial van één minuut bestaan ligt in de orde van grootte van 0,6, zoals bleek uit de intercorrelaties tussen de afzonderlijke letters. Dit is onvoldoende voor gebruik in de diagnostiek van individuen. Men dient dus in ieder geval een variant met ten minste drie letters af te nemen. Wil men meten met een hogere betrouwbaarheid dan 0,8, dan moet men nog meer trials afnemen.

Wat betreft de constructvaliditeit is van belang dat de prestatie op de letterfluencytest behalve met de woordenschat ook vrij sterk samenhangt met categoriefluency, veel minder met non-verbale fluency (figuurfluency), en slechts matig met een andere executieve functie test (Stroop kleur-woord test). De betrekkelijk sterke samenhang met woordenschat en de lage correlatie met figuurfluency is ook over de COWAT gemeld.<sup>41</sup> Dat de test niet correleert met benoemen (Boston Naming Test) komt waarschijnlijk doordat de BNT weinig differentieert in een normale populatie.<sup>32</sup>

Sommige auteurs menen dat de productie aan het begin van elke trial vooral een beroep doet op automatische processen van informatieverwerking, terwijl daarna de woordproductie tot stand komt door meer gecontroleerde zoekprocessen in het lexicale en semantische geheugen.<sup>27,28</sup> Onze resultaten steunen die gedachte niet. We vonden dat de invloed zowel van woordkennis, of wellicht verbale intelligentie (gemeten met de NLV), als van gecontroleerde informatieverwerking (selectieve aandacht en inhibitie gemeten met de Stroop taak) constant blijven tijdens de hele taak. We konden dus de bevindingen van Fernaeus en Almkvist niet repliceren.<sup>27</sup> Zij deden hun onderzoek bij oudere patiënten in een geheugenpolikliniek, terwijl wij gezonde



personen onderzochten. Waarschijnlijk liggen de relaties tussen fluency en andere aspecten van het executief functioneren in patiëntenpopulaties anders dan in de normale bevolking.

In de inleiding hebben we verwezen naar literatuur over de klinische toepassing van de test bij diverse neurologische en psychiatrische aandoeningen. Het is niet onze bedoeling hier diep op dit onderwerp in te gaan. Eén klinische toepassing willen we echter niet onvermeld laten, omdat die goed illustreert dat de test responsief is, een ander belangrijk gegeven voor de validiteit. Dit betreft neuropsychologisch onderzoek naar eventuele nadelige effecten van neurochirurgische behandeling van patiënten met de ziekte van Parkinson. Letterfluency blijkt één van de gevoeligste detectoren van cognitieve veranderingen na pallido- en thalamotomie en deep brain stimulation te zijn.<sup>42-44</sup> Vooral als de ingreep in de linker hemisfeer wordt gedaan, is het effect op de verbale fluency aanzienlijk.<sup>45-47</sup>

Zwakke punten van deze normeringsstudie zijn het beperkte aantal proefpersonen en de gebrekkige representativiteit van de steekproef. De groep bestond uit 200 personen, een aantal dat over het algemeen beschouwd wordt als te klein voor een goede normering.<sup>48</sup> Dit bezwaar wordt minder relevant als men in aanmerking neemt dat er geen belangrijke effecten van leeftijd en geslacht zijn. Het is daarom niet nodig de

steekproef op te delen in subgroepen. Wel is er een effect van het opleidingsniveau, en juist op dit punt is de steekproef niet representatief voor de bevolking. Om deze reden hebben we een regressieformule aangeleverd waarmee we tabel 3 hebben opgesteld, zodat de score op de test gecorrigeerd kan worden voor het opleidingsniveau (zie ook Appendix B). Daarmee wordt enigszins tegemoet gekomen aan het bezwaar van beperkte representativiteit van de steekproef. Bij toepassing van deze formule vergelijkt men immers in feite de score met die van personen met hetzelfde niveau van opleiding.

Ondanks deze beperkingen menen wij dat dit een goede Nederlandstalige versie is van de internationaal meest gebruikte letterfluencytest.

#### Dankwoord

Met dank aan Hilde Geurts en Harriët Smeding voor hun commentaar op een eerdere versie.

[1] Tijdens de reviewprocedure werden wij geattendeerd op de publicatie van P.C. Rolf en R.J. van Rijnsoever (Positionele letterfrequenties van het Nederlands. Lisse, Swets & Zeitlinger, 1984). Deze publicatie bevat tellingen van het aantal woorden die in het Nederlands beginnen met een A, een B enz. en is evenals de door ons gebruikte methode een geschikte basis voor een selectie van letters in een letterfluencytaak.

#### Literatuur

- 1 Thurstone LL. Primary mental abilities. Chicago: Chicago University Press; 1938.
- 2 Guilford JP. Three faces of intellect. *American Psychologist*. 1959;14:469-479.
- 3 Hudson PTW. Preliminary category norms for verbal items in 51 categories in Dutch. Nijmegen: Katholieke Universiteit; 1982.
- 4 Battig WF, Montague WE. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph* 1969;80:1-46.
- 5 Van Overschelde JP, Rawson KA, Dunlosky J. Category norms: An updated and expanded version of the Battig and Montague (1969) norms (vol 50, pg 289, 2004). *Journal of Memory and Language* 2006;54(4):633-633.
- 6 Mulder JL, Dekker PH, Dekker R. *Figuur-Fluency Test*. Leiden: PITS; 2006.
- 7 Jones-Gotman M, Milner B. Design fluency: the invention of nonsense drawings after focal cortical lesions. *Neuropsychologia* 1977;15(4-5):653-74.
- 8 Ruff RM. *Ruff Figural Fluency Test administration manual*. San Francisco: Neuropsychological Resources; 1988.
- 9 Cabeza R, Nyberg L. Imaging cognition II: An empirical review of 275 PET and fMRI studies. *J Cogn Neurosci* 2000;12(1):1-47.
- 10 Henry JD, Crawford JR. A meta-analytic review of verbal fluency deficits in depression. *J Clin Exp Neuropsychol* 2005;27(1):78-101.
- 11 Henry JD, Crawford JR. A meta-analytic review of verbal fluency deficits in schizophrenia relative to other neurocognitive deficits. *Cognit Neuropsychiatry* 2005;10(1):1-33.
- 12 Henry JD, Crawford JR. A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. *Neuropsychology* 2004;18(4):621-8.
- 13 Henry JD, Crawford JR. Verbal fluency deficits in Parkinson's disease: a meta-analysis. *J Int Neuropsychol Soc* 2004;10(4):608-22.
- 14 Henry JD, Crawford JR. A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology* 2004;18(2):284-95.
- 15 Henry JD, Crawford JR, Phillips LH. A meta-analytic review of verbal fluency deficits in Huntington's disease. *Neuropsychology* 2005;19(2):243-52.
- 16 Henry JD, Crawford JR, Phillips LH. Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia* 2004;42(9):1212-22.

- 17 Lezak MD, Howieson DB, Loring DW. *Neuropsychological Assessment*. Fourth ed. New York: Oxford University Press; 2004.
- 18 Bokat CE, Goldberg TE. Letter and category fluency in schizophrenic patients: a meta-analysis. *Schizophr Res* 2003;64(1):73-8.
- 19 Cabeza R, Nyberg L. Imaging cognition: An empirical review of PET studies with normal subjects. *Journal of Cognitive Neuroscience* 1997;9(1):1-26.
- 20 Troyer AK. Normative data for clustering and switching on verbal fluency tasks. *J Clin Exp Neuropsychol* 2000;22(3):370-8.
- 21 Audenaert K, Brans B, Van Laere K, Lahorte P, Versijpt J, van Heeringen K, et al. Verbal fluency as a prefrontal activation probe: a validation study using 99mTc-ECD brain SPET. *Eur J Nucl Med* 2000;27(12):1800-8.
- 22 Gazzaniga MS, Ivry RB, Mangun GR. *Cognitive neuroscience. The biology of mind*. 2nd ed. New York: W.W. Norton & Company; 2002.
- 23 Luteijn F, Barelds DPF. *Groninger Intelligentie Test-2 (GIT-2)*. Amsterdam: Harcourt Test Publishers; 2004.
- 24 Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. Normative data for the Animal, Profession and Letter M Naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. *J Int Neuropsychol Soc* 2006;12(1):80-9.
- 25 Lannoo E, Vingerhoets G. Flemish normative data on common neuropsychological tests: influence of age, education, and gender. *Psychologica Belgica* 1997;37:141-155.
- 26 Benton AL, Hamsher K. *Multilingual Aphasia Examination*. Iowa City: AJA Associates; 1989.
- 27 Fernaeus SE, Almkvist O. Word production: Dissociation of two retrieval modes of semantic memory across time. *Journal of Clinical and Experimental Neuropsychology* 1998;20(2):137-143.
- 28 Hurks PPM, Vles JSH, Hendriksen JGM, Kalff AC, Feron FJM, Kroes M, et al. Semantic category fluency versus initial letter fluency over 60 seconds as a measure of automatic and controlled processing in healthy school-aged children. *Journal of Clinical and Experimental Neuropsychology* 2006;28(5):684-695.
- 29 CBS. *Statistisch Jaarboek 2003*. Voorburg/Heerlen: Centraal Bureau voor de Statistiek; 2003.
- 30 Kaplan E, Goodglass H, Weintraub S. *The Boston Naming Test*. Philadelphia: Lea and Febiger; 1983.
- 31 Marien P, Mampaey E, Vervaeke A, Saerens J, De Deyn PP. Normative data for the Boston Naming Test in native Dutch-speaking Belgian elderly. *Brain and Language* 1998;65(3):447-467.
- 32 Hammes JGW. *De Stroop Kleur-woord test*. Lisse: Swets en Zeitlinger; 1978.
- 33 Schmand B, Houx P, de Koning I. Normen voor Stroop kleur-woord tests, Trail Making test, en Story Recall van de Rivermead Behavioural Memory Test. Amsterdam: Nederlands Instituut van Psychologen; 2004. [www.neuropsycholoog.nl](http://www.neuropsycholoog.nl).
- 34 Schmand B, Lindeboom J, Van Harskamp F. *Nederlandse Leestest voor volwassenen*. Lisse: Swets & Zeitlinger; 1992.
- 35 Solomon PR, Sullivan DM, Pendlebury WW. Toward recognition of the Alzheimer's disease patient in primary care practice: The 7 minute screen. *Neurology* 1998;50(4):A162-A162.
- 36 Meulen EFJ, Schmand B, van Campen JP, de Koning SJ, Ponds RW, Scheltens P, et al. The seven minute screen: a neurocognitive screening test highly sensitive to various types of dementia. *J Neurol Neurosurg Psychiatry* 2004;75(5):700-705.
- 37 Benton AL, Hamsher K, Sivan AB. *Multilingual Aphasia Examination*, 3rd edition. Iowa City: AJA Associates; 1994.
- 38 Strauss E, Sherman EMS, Spreen O. *A compendium of neuropsychological tests. Administration, norms, and commentary*. Third edition. Third ed. New York: Oxford University Press; 2006.
- 39 Loonstra AS, Tarlow AR, Sellers AH. COWAT metanorms across age, education, and gender. *Appl Neuropsychol* 2001;8(3):161-6.
- 40 Rodriguez-Aranda C, Martinussen M. Age-related differences in performance of phonemic verbal fluency measured by Controlled Oral Word Association Task (COWAT): a meta-analytic study. *Dev Neuropsychol* 2006;30(2):697-717.
- 41 Ruff RM, Light RH, Parker SB, Levin HS. The psychological construct of word fluency. *Brain and Language* 1997;57(3):394-405.
- 42 Woods SP, Fields JA, Troster AI. Neuropsychological sequelae of subthalamic nucleus deep brain stimulation in Parkinson's disease: A critical review. *Neuropsychology Review* 2002;12(2):111-126.
- 43 Lacritz LH, Cullum CM, Frol AB, Dewey RB, Jr., Giller CA. Neuropsychological outcome following unilateral stereotactic pallidotomy in intractable Parkinson's disease. *Brain Cogn* 2000;42(3):364-78.
- 44 Voon V, Kubu C, Krack P, Houeto JL, Troster AI. Deep brain stimulation: neuropsychological and neuropsychiatric issues. *Mov Disord* 2006;21 Suppl 14:S305-27.
- 45 Schmand B, de Bie RM, Koning-Haanstra M, de Smet JS, Speelman JD, van Zomeren AH. Unilateral pallidotomy in PD: a controlled study of cognitive and behavioral effects. The Netherlands Pallidotomy Study (NEPAS) group. *Neurology* 2000;54(5):1058-64.
- 46 Smeding HM, Speelman JD, Koning-Haanstra M, Schuurman PR, Nijssen P, van Laar T, et al. Neuropsychological effects of bilateral STN stimulation in Parkinson disease: a controlled study. *Neurology* 2006;66(12):1830-6.
- 47 Smeding HM, Esselink RAJ, Schmand B, Koning-Haanstra M, Nijhuis I, Wijnalda EM, et al. Unilateral pallidotomy versus bilateral subthalamic nucleus stimulation in PD - A com-

- parison of neuropsychological effects. *J Neurol* 2005;252(2):176-182.
- 48 Evers A, van Vliet-Mulder JC, Groot CJ. Documentatie van Tests en Testresearch in Nederland, deel 1 en 2. Amsterdam/Assen: NIP/Van Gorcum; 2000.
- 49 Verhage F. Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar. Assen: Van Gorcum; 1964.
- 50 Hendriks M, Kessels R, Gorissen M, Schmand B. Neuropsychologische diagnostiek. De klinische praktijk. Amsterdam: Boom; 2006.
- 51 Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991;59(1):12-9.
- 52 Chelune G-J, Naugle R-I, Luders H, Sedlak J, et al. Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology* 1993;7(1):41-52.
- 53 Peters, M. J. V, Ponds, R. W. H. M, van de Ham, P, Scheltens, Ph, Verhey, F. R. J. Detectie van dementie van het Alzheimer type (DAT) met de '7 minuten test' (7 MT). *Tijdschr Gerontol Geriatr* 2004; 35: 114-120.

# Bijlage

## Appendix A. Instructie letterfluency

De opdracht is om in één minuut zoveel mogelijk verschillende woorden te noemen die met een bepaalde letter beginnen. Eigennamen gelden niet, evenmin als getallen of cijfers.

Zie voor volledige instructie: [www.tgg.bsl.nl](http://www.tgg.bsl.nl).

## Appendix B. Normen letterfluency

Het verwachte aantal woorden op grond van het opleidingsniveau wordt berekend met de onderstaande regressieformule. Deze werd verkregen m.b.v. een lineaire regressie-analyse met het opleidingsniveau als onafhankelijke variabele. De standaardfout van de schatting is 10 woorden, en de proportie verklaarde variantie is 8% (adjusted R-square).

De residuscores zijn normaal verdeeld.

Verwacht aantal woorden =  $25 + 2,5 \times \text{opleiding}$

Het feitelijk opgenoemde aantal woorden kan vervolgens worden uitgedrukt in een voor opleidingsniveau gecorrigeerde T-score (gemiddelde = 50, standaarddeviatie = 10) met behulp van de volgende formule.

T-score =  $50 + (\text{aantal opgenoemde woorden} - \text{verwacht aantal woorden}) / \text{standaardfout}$

Door substitutie van het verwacht aantal woorden kan deze formule worden vereenvoudigd:

T-score =  $\text{aantal opgenoemde woorden} + 25 - 2,5 \times \text{opleiding}$

Het opleidingsniveau wordt gescoord volgens het systeem van Verhage.<sup>49</sup>

Zie Hendriks et al. pagina 122 voor een revisie van dit systeem.<sup>50</sup> In het kort komt dit neer op het volgende.

- 1 = niet voltooid basisonderwijs
- 2 = basisonderwijs
- 3 = niet voltooid vervolgonderwijs
- 4 = lbo, vmbo-beroepsgericht
- 5 = mbo, mulo, mavo, vmbo-theoretisch
- 6 = havo, vwo, hbo
- 7 = universitair onderwijs

Voorbeeld: een patiënt met een lagere beroepsopleiding behaalt een score van 22 op de letterfluency-test. Zijn verwachte score is  $25 + 2,5 \times 4 = 35$  woorden. Hij presteert dus minder goed dan verwacht. Zijn T-score is  $22 + 25 - 2,5 \times 4 = 37$ . Dit is 1,3 SD onder het gemiddelde van zijn opleidingsniveau, en komt overeen met het 10e percentiel.

Het verwachte aantal woorden kan ook worden geschat met behulp van de NLV<sup>(34)</sup>. De NLV verklaart iets meer variantie (13%; adjusted R-square) dan de opleiding. De standaardfout van de schatting is 10 woorden. De residuscores zijn normaal verdeeld.

Verwacht aantal woorden =  $0,47 \times \text{NLV ruwe score} - 2$

T-score =  $\text{aantal opgenoemde woorden} + 52 - 0,47 \times \text{NLV ruwe score}$

## Appendix C. Kritische verschillscore bij herhaalde metingen

Onder de aanname dat de hertestbetrouwbaarheid 0,80 en de standaard deviatie 10,5 is, bedraagt de standaard meetfout 4,7 en is de standaardfout van het verschil 6,6 (reliable change index)<sup>(51)</sup>. De kritische verschillscore is dan 11 woorden ( $p=0,05$  eenzijdig). Dus de prestatie is significant beter wanneer

er bij een tweede meting ten minste 12 woorden meer worden geproduceerd dan de eerste keer, althans wanneer we mogen aannemen dat er geen hertesteffect is.

Bij cognitief intacte personen, en bij patiënten bij wie nog een normaal hertesteffect te verwachten is, moet voor het hertesteffect van 4 woorden worden gecorrigeerd (<sup>52</sup>). Men dient dan wel een andere versie te gebruiken dan de eerste keer. De kritische verschilscore is dan 15 woorden (11+4). Onder die omstandigheden kan van een significante verslechtering worden gesproken bij een scorevermindering van meer dan 7 woorden (11-4).

De kritische verschilscore is 15,5 woorden bij eenzijdige toetsing op een significantieniveau van  $p=0,01$ . Men moet dus ten minste 16 woorden meer produceren wil de scoreverbetering significant op 1% genoemd kunnen worden (en tenminste 20 woorden meer als rekening gehouden moet worden met het hertesteffect).