

De Akense Afasie Test (2)

R. De Bleser / K. Willmes / P. Graetz / P. Hagoort

In het eerste artikel (Graetz e.a., 1991) zijn de eigenschappen van de constructie van de Akense Afasie Test (AAT; Huber e.a., 1983) alsmede de aanpassing aan de regulariteiten van de Nederlandse taal beschreven. Dit tweede artikel bericht over de standaardisatieprocedure van de Nederlandstalige AAT waarin verschillende aspecten van validiteit en betrouwbaarheid zijn onderzocht bij een steekproef van 120 afasiepatiënten en 60 controleproefpersonen. Het blijkt dat de Nederlandse AAT valide is met betrekking tot de toenemende complexiteit van opdrachten binnen opdrachtgroepen en van opdrachtgroepen binnen subtests. De AAT bezit daarnaast een hoge differentiële validiteit ten aanzien van zowel het onderscheid tussen afasie en niet-afasie alsmede tussen afasiesyndromen. De hoge interne consistentie van de subtests maakt verantwoorde individuele diagnostiek mogelijk.

Inleiding

Bij de constructie van de Nederlandstalige versie van de Akense Afasie Test (AAT) zijn zoveel mogelijk dezelfde linguïstische criteria gehanteerd als bij de oorspronkelijke Duitse versie. In dit tweede artikel worden de psychometrische eigenschappen objectiviteit, validiteit en betrouwbaarheid beschreven. Alleen wanneer die bevredigend uitvallen, kunnen de in het eerste artikel beschreven doelstellingen van de AAT verwezenlijkt worden:

- (1) scheiding tussen patiënten met en zonder afasie;
- (2) vaststelling van de ernst van de afatische taalstoornissen in verschillende taalmodaliteiten;
- (3) classificatie van de taalstoornissen;
- (4) beoordeling van het verloop van de afasie bij individuele patiënten.

Het derde en laatste artikel in deze serie zal beschrijven hoe de testgebruiker in de praktijk met de AAT deze doelen kan bereiken.

Objectiviteit

Teneinde de *afname* van de AAT zo objectief mogelijk te houden is, net als voor de Duitse versie, in de handleiding (Graetz e.a., in voorbereiding) precies vastgelegd hoe de diverse onderdelen van de AAT afgenomen moeten worden.

Met uitzondering van de Token Test, waarvan de antwoorden in termen van goed/fout worden beoordeeld, worden alle expressieve en receptieve talige reacties gedifferentieerd beoordeeld (0-5 voor de beoordelingsschalen van de spontane taalproductie

en 0-3 voor de opdrachten van de subtests). Daarom dient empirisch te worden getoetst, in hoeverre verschillende beoordelaars bij dezelfde patiënten tot overeenkomende scores komen (*tussenbeoordelaarsbetrouwbaarheid*).

Zoals door Weniger e.a. (1981) en in Huber e.a. (1983) is beschreven, zijn de resultaten van een onderzoek naar deze tussenbeoordelaarsbetrouwbaarheid voor de Duitse versie zeer bevredigend. In dit onderzoek hebben 29 studenten van de Lehranstalt für Logopädie in Aken en 20 testgebruikers in verschillende neurologische klinieken in Duitsland 48 AAT's opnieuw beoordeeld. Voor de zes beoordelingsniveaus van de spontane taalproductie resulteerden deze scoringen in overeenstemmingscoëfficiënten (Krippendorff, 1970) tussen 0.75 en 0.86. De somwaarden van de beoordelingen van de subtests met expressieve talige uitingen leverden coëfficiënten van rond de 0.99 op. Aangezien de Nederlandstalige versie van de AAT volledig identieke beoordelingscriteria hanteert, is er geen aanleiding om afwijkende resultaten te verwachten. Daarom is afgezien van een herhaald onderzoek van de tussenbeoordelaarsbetrouwbaarheid bij Nederlandstalige testgebruikers.

Validiteit

Aan de validiteit van de test zijn diverse aspecten te onderscheiden. In de eerste plaats moeten met het oog op *constructvaliditeit* de constructieprincipes empirisch worden onderzocht. Als eerste rijst de vraag, of de linguïstisch gemotiveerde combinatie van opdrachten in opdrachtgroepen en van deze opdrachtgroepen in subtests gerechtvaardigd is. Vervolgens moet het constructieprincipe van de toenemende moeilijkheidsgraad van opdrachten binnen een opdrachtgroep alsmede van opdrachtgroepen binnen een subtest empirisch worden aangetoond. *Differentiële validiteit* bezit de test, wanneer afatische patiënten ermee van niet-afatische patiënten onderscheiden kunnen worden en er bovendien een goede differentiatie tussen de verschillende afatische syndromen mogelijk is.

Betrouwbaarheid

Eenzijds dient voor de opdrachtgroepen en afzonderlijke subtests te worden aangetoond dat ze intern *consistent* zijn. Dit betekent dat de scores onafhankelijk van de betreffende opdrachten worden behaald: de opdrachten stellen dezelfde talige eisen. Hiervan is sprake wanneer de onderzochte patiënten door de behaalde scores in steeds (ongeveer)

dezelfde volgorde qua ernst van de stoornis gerangschikt worden. Anderzijds moeten de talige prestaties bij herhaling van de test verregaand reproduceerbaar zijn (*test-hertestbetrouwbaarheid*). Voor de Duitstalige versie van de AAT is aangetoond (Weniger e.a., 1981) dat bij herhaling van de test na slechts twee dagen geen oefen- of geheugeneffecten optreden en dat de prestaties op de subtests zeer hoog met elkaar correleren. Er is wederom geen reden te veronderstellen dat zulke effecten bij de Nederlandstalige versie van de AAT wél zouden optreden, noch dat de talige prestaties van Nederlandse patiënten sterker zouden fluctueren. Daarom is geen nieuw onderzoek naar de test-hertestbetrouwbaarheid uitgevoerd.

Steekproeven

Sinds 1982 zijn in een dertigtal instellingen in Nederland grote aantallen afasie- en controlepatiënten met de Nederlandstalige AAT onderzocht. Op grond van regionale taalverschillen was het nodig om mensen uit uiteenlopende dialectgebieden te onderzoeken. Tot de instellingen waar patiënten met de AAT zijn onderzocht behoren neurologische afdelingen van ziekenhuizen, revalidatiecentra en verpleeghuizen (zie het dankwoord, Graetz e.a., 1991).

Al deze AAT-onderzoeken zijn door de Afasiegroep van het Max-Planck-Instituut te Nijmegen verzameld ten behoeve van zowel de standaardisatie- alsmede de normeringsprocedure. In dit artikel worden de resultaten van de standaardisatiefase gerapporteerd. Voor opname in de standaardisatiesteekproef golden een aantal criteria: leeftijd tot 75 jaar, Nederlands als moedertaal, oorspronkelijk rechtshandig, linker hersenhelftbeschadiging door infarct of bloeding en in uitzonderingsgevallen door een trauma; deze moest bovendien door een CT-scan zijn bevestigd.

Net als bij de Duitstalige AAT omvat de Nederlandse standaardisatiesteekproef zes groepen van elk 30 mensen: één groep voor elk van de vier afasische hoofdsyndromen, een groep met 30 controlepatiënten met rechter hersenhelftbeschadiging zonder afasie en een groep met 30 controleproefpersonen zonder hersenbeschadiging. Uitsluitend aan de hand van het semi-gestandaardiseerde interview is elke patiënt door drie ervaren beoordelaars geclassificeerd; zij hebben hun oordelen onafhankelijk van elkaar uitgesproken. De gehanteerde classificatiecriteria zijn de syndroombeschrijvingen van Poeck e.a. (1975). Deze expert-classificaties fungeren als extern criterium bij de toetsing van de differentiële validiteit van de AAT. Natuurlijk is het niet mogelijk om elke patiënt op basis van het interview unaniem aan één van de vier hoofdsyndroomgroepen (globale, Wernicke, Broca of amnestische afasie) toe te kennen; daarnaast hebben de beoordelaars sommige patiënten niet eensluidend kunnen classificeren en een aantal andere patiënten als niet-classificeerbaar, rest-afasisch of niet-afasisch gecategoriseerd. Al deze testafnames zijn opgenomen in de normeringssteekproef (zie Willmes e.a., in voorbereiding).

Om een zo groot mogelijke vergelijkbaarheid met de oorspronkelijke Duitse steekproef te verwezenlijken, zijn niet simpelweg de eerste 30 afasische patiënten per hoofdsyndroomgroep opgenomen die aan de bovenstaande criteria voldoen. In plaats daarvan zijn per groep 30 patiënten geselecteerd, van wie het type afasie, de leeftijd en het geslacht zoveel mogelijk met de afasiepatiënten in de Duitse standaardisatiesteekproef overeenkomen. De duur van de afasie varieert van 1 tot 66 maanden. Een deel van de patiënten verbleef nog in het ziekenhuis, terwijl een ander deel thuis met de AAT is onderzocht. Voor de controlegroep van 30 mensen zonder hersenbeschadiging zijn daarom deels mensen met niet-neurologische aandoeningen in ziekenhuizen of revalidatiecentra met de AAT getest, en deels mensen thuis, bijvoorbeeld echtgenoten van afasiepatiënten. In tabel 1 zijn de eigenschappen van de proefpersoongroepen weergegeven.

Groep (n=30)	Geslacht		Leeftijd (j)		Duur (m)		Etiologie	
	v	m	Md	bereik	Md	bereik	CVA	overig
Globaal	12	18	60	30-72	7	1-66	30	–
Wernicke	13	17	62	28-74	4	1-24	30	–
Broca	10	20	55	26-76	4.5	1-65	28	2
Amnestisch	9	21	58	25-70	3	1-42	28	2
Re-hemisfeer	14	16	55.5	21-75	3	1-60	30	–
Normaal	18	12	56	30-71	–	–	–	–

Tabel 1. Kenmerken van de standaardisatiesteekproef.

De Akense Afasie Test

Voor een helder begrip van de resultaten volgt hier een korte samenvatting van de opbouw van de AAT. De geïntroduceerde afkortingen voor spontane taalproductie en subtests worden in de tabellen en afbeeldingen gebruikt. De AAT bestaat uit een beoordeling van de spontane taalproductie en vijf subtests: Token Test in de versie met 50 opdrachten volgens Orgass (1986), Naspreken, Schrijftaal, Benoemen en auditief en leesinhoudelijk Taalbegrip (respectievelijk TT, NA, ST, BE en TB).

De uitingen ter beoordeling van de spontane taalproductie worden in een semi-gestandaardiseerd interview over vertrouwde thema's verzameld. Op zes ordinale schalen met elk zes mogelijke puntwaarden (0-5) worden vervolgens het communicatief gedrag, de articulatie en prosodie, het geautomatiseerd taalgebruik en de semantische, fonematische en syntactische structuur afzonderlijk beoordeeld (respectievelijk COM, ART, AUT, SEM, FON, SYN).

De vijf subtests van de AAT bestaan uit drie (ST), vier (BE, TB) of vijf (TT, NA) groepen van elk tien opdrachten. Met uitzondering van de Token Test worden de antwoorden op een vierpunts-schaal met puntwaarden 0-3 beoordeeld. Hierbij betekent 0 het ontbreken van een respons of een zeer zware afwijking, 1 een sterke, 2 een lichte en 3 geen afwijking van de doelvorm.

	Alle afatici (n=120)		Globaal (n=30)		Wernicke (n=30)		Broca (n=30)		Amnestisch (n=30)		Re-Hemisfeer (n=30)		Normaal (n=30)	
SPONTANE TAALPRODUKTIE														
COM	2.0	(0-5)	0.0	(0-1)	2.5	(0-4)	2.0	(1-3)	3.0	(2-5)	5.0	(4-5)	5.0	(5)
	2.01	(1.31)	0.4	(0.5)	2.2	(1.3)	2.1	(0.6)	3.3	(0.6)	4.7	(0.5)	5.0	(0.0)
ART	4.0	(0-5)	1.0	(0-5)	5.0	(3-5)	3.0	(2-5)	5.0	(2-5)	5.0	(3-5)	5.0	(5)
	3.48	(1.65)	1.7	(1.8)	4.7	(0.5)	3.0	(0.9)	4.5	(0.8)	4.7	(0.6)	5.0	(0.0)
AUT	3.0	(0-5)	0.0	(0-3)	2.5	(2-5)	3.0	(2-5)	4.0	(2-5)	5.0	(3-5)	5.0	(3-5)
	2.62	(1.54)	0.6	(0.7)	2.8	(0.9)	3.2	(1.0)	3.9	(0.9)	4.6	(0.6)	4.9	(0.4)
SEM	3.0	(0-4)	0.0	(0-1)	3.0	(0-4)	3.0	(2-4)	3.0	(3-4)	5.0	(4-5)	5.0	(4-5)
	2.28	(1.43)	0.2	(0.4)	2.3	(1.0)	3.3	(0.5)	3.3	(0.5)	4.7	(0.5)	4.9	(0.3)
FON	3.0	(0-5)	0.0	(0-2)	4.0	(1-4)	3.0	(2-5)	4.0	(3-5)	5.0	(3-5)	5.0	(4-5)
	2.67	(1.67)	0.2	(0.5)	3.2	(0.9)	3.2	(1.0)	4.1	(0.6)	4.6	(0.6)	4.8	(0.4)
SYN	2.5	(0-5)	0.0	(0-1)	3.0	(0-3)	2.0	(1-4)	3.0	(3-5)	4.0	(4-5)	5.0	(4-5)
	2.04	(1.39)	0.1	(0.3)	2.9	(0.6)	1.8	(0.6)	3.4	(0.6)	4.4	(0.5)	4.8	(0.4)
TOKEN TEST	36.0	(0-50)	48.0	(20-50)	42.0	(3-50)	34.0	(1-50)	14.5	(0-41)	1.0	(0-10)	1.0	(0-3)
(fouten)	31.9	(15.6)	44.5	(8.1)	36.3	(15.2)	29.0	(13.6)	17.6	(11.0)	2.2	(3.0)	0.9	(1.0)
1-4	26.0	(0-40)	38.0	(13-40)	32.0	(0-40)	24.0	(0-40)	10.0	(0-31)	0.0	(0-7)	0.0	(0-2)
	24.1	(13.0)	34.7	(7.5)	27.7	(13.1)	21.3	(11.0)	12.6	(8.3)	1.0	(1.9)	0.5	(0.7)
5	10.0	(0-10)	10.0	(7-10)	10.0	(3-10)	10.0	(0-10)	4.5	(0-10)	1.0	(0-4)	0.0	(0-2)
	7.8	(3.0)	9.8	(0.7)	8.7	(2.3)	7.7	(3.1)	5.0	(3.1)	1.2	(1.3)	0.4	(0.7)
NASPREKEN	94.5	(0-50)	33.0	(0-114)	82.0	(10-147)	105.0	(28-129)	141.0	(77-150)	146.5	(130-150)	149.0	(143-150)
	87.9	(46.2)	40.0	(31.2)	78.6	(46.6)	99.3	(23.1)	133.8	(18.1)	144.3	(5.5)	148.3	(1.8)
NA1	25.5	(0-30)	10.5	(0-27)	28.0	(4-30)	26.0	(6-30)	29.0	(20-30)	30.0	(23-30)	30.0	(26-30)
	22.0	(9.1)	13.3	(9.0)	22.8	(9.0)	24.1	(6.6)	28.1	(2.6)	28.9	(1.9)	29.7	(0.9)
NA2	25.0	(0-30)	9.5	(0-29)	23.5	(2-30)	25.5	(8-30)	30.0	(25-30)	29.0	(26-30)	30.0	(27-30)
	21.1	(9.8)	11.3	(9.2)	19.5	(10.1)	24.8	(4.6)	29.1	(1.3)	29.1	(1.0)	29.7	(0.6)
NA3	21.0	(0-30)	6.0	(0-26)	16.5	(0-30)	22.0	(11-29)	29.0	(18-30)	30.0	(22-30)	30.0	(28-30)
	18.3	(10.2)	8.3	(7.6)	15.7	(11.0)	21.1	(5.1)	27.9	(3.0)	29.2	(1.8)	29.7	(0.6)
NA4	13.5	(0-30)	1.5	(0-22)	7.0	(0-30)	18.0	(0-28)	27.5	(1-30)	29.0	(24-30)	30.0	(27-30)
	14.2	(10.8)	4.8	(6.4)	10.7	(10.5)	16.2	(6.6)	25.0	(7.1)	28.6	(1.8)	29.4	(0.7)
NA5	10.5	(0-30)	1.0	(0-16)	6.5	(0-30)	13.0	(0-26)	26.0	(4-30)	29.0	(23-30)	30.0	(28-30)
	12.4	(10.5)	2.9	(4.1)	9.9	(10.1)	13.1	(7.3)	23.7	(7.1)	28.4	(2.0)	29.6	(0.6)
SCHRIJFTAAL	41.5	(0-90)	3.0	(0-50)	33.0	(0-85)	42.0	(5-85)	81.5	(14-90)	87.5	(59-90)	89.0	(86-90)
	42.3	(31.5)	9.6	(13.6)	38.2	(28.8)	45.6	(22.6)	75.7	(16.2)	85.2	(6.3)	88.8	(1.4)
ST1	19.5	(0-30)	0.0	(0-24)	15.5	(0-30)	20.5	(5-29)	29.0	(3-30)	29.0	(20-30)	30.0	(28-30)
	16.5	(11.3)	3.5	(6.0)	16.4	(11.2)	18.9	(6.4)	27.1	(5.3)	28.5	(2.2)	29.8	(0.5)
ST2	13.5	(0-30)	0.0	(0-20)	12.0	(0-30)	14.5	(0-30)	27.0	(0-30)	29.0	(18-30)	30.0	(28-30)
	14.1	(10.9)	3.9	(5.7)	1.2	(10.1)	15.4	(9.1)	24.7	(6.5)	28.5	(2.3)	29.5	(0.7)
ST3	10.5	(0-30)	0.0	(0-15)	8.5	(0-27)	8.0	(0-29)	26.5	(9-30)	29.0	(20-30)	30.0	(28-30)
	11.8	(11.1)	2.2	(4.3)	9.6	(9.7)	11.3	(10.0)	23.9	(6.4)	28.1	(2.6)	29.5	(0.7)
BENOEMEN	67.5	(0-115)	1.5	(0-47)	68.0	(0-110)	79.0	(20-112)	91.0	(28-115)	108.0	(95-116)	115.0	(108-119)
	58.1	(38.2)	11.7	(16.1)	60.3	(35.8)	73.9	(24.0)	86.6	(23.4)	107.7	(4.8)	115.2	(2.6)
BE1	22.0	(0-30)	0.0	(0-24)	20.5	(0-30)	24.5	(3-30)	25.5	(2-30)	28.0	(19-30)	29.0	(23-30)
	16.9	(11.1)	5.0	(7.1)	17.2	(10.9)	22.2	(7.1)	23.2	(8.1)	27.8	(2.5)	28.8	(1.6)
BE2	20.5	(0-30)	0.0	(0-19)	21.5	(0-30)	22.0	(7-30)	24.5	(6-30)	27.5	(21-30)	29.0	(24-30)
	16.6	(10.5)	4.1	(6.4)	17.3	(10.4)	21.2	(6.2)	23.8	(5.2)	27.1	(2.2)	28.7	(1.4)
BE3	14.0	(0-30)	0.0	(0-17)	12.0	(0-28)	17.5	(0-29)	21.0	(0-30)	29.0	(25-30)	29.0	(27-30)
	12.9	(10.5)	1.7	(3.5)	13.0	(9.7)	16.9	(8.4)	20.1	(8.5)	28.2	(1.8)	29.1	(1.0)
BE4	12.5	(0-28)	0.0	(0-9)	12.5	(0-27)	13.5	(0-26)	21.0	(5-28)	25.0	(17-29)	29.0	(25-30)
	11.7	(9.0)	0.9	(2.0)	12.7	(7.9)	13.6	(6.8)	19.5	(5.4)	24.5	(3.0)	28.7	(1.5)
TAALBEGRIIP	77.5	(0-114)	62.5	(0-94)	73.0	(27-113)	83.5	(55-114)	92.5	(67-114)	106.0	(84-119)	115.0	(105-120)
	77.6	(21.1)	59.6	(18.8)	74.0	(21.1)	84.3	(14.9)	92.6	(13.4)	105.2	(9.4)	114.6	(3.6)
TB1	20.0	(0-30)	19.0	(0-28)	18.5	(10-27)	21.5	(14-30)	22.5	(13-30)	25.5	(18-30)	29.0	(23-30)
	20.3	(5.2)	17.8	(6.3)	19.2	(4.2)	21.9	(4.4)	22.5	(4.6)	25.2	(3.2)	28.4	(2.0)
TB2	21.0	(0-30)	14.5	(0-28)	20.5	(7-30)	22.0	(12-28)	24.5	(13-30)	26.5	(19-30)	30.0	(23-30)
	20.1	(5.9)	15.6	(5.5)	18.8	(5.7)	21.8	(4.7)	24.0	(4.0)	25.8	(3.2)	28.6	(1.9)
TB3	22.0	(0-30)	16.0	(0-28)	22.0	(0-30)	22.0	(15-30)	24.5	(14-30)	28.0	(20-30)	30.0	(27-30)
	20.8	(6.8)	15.3	(6.1)	20.5	(8.2)	22.6	(4.5)	24.8	(4.1)	28.0	(2.3)	29.2	(1.1)
TB4	17.0	(0-30)	11.5	(0-23)	15.0	(0-30)	18.0	(1-28)	24.0	(7-30)	27.0	(17-30)	30.0	(20-30)
	16.4	(7.4)	11.0	(5.8)	15.5	(8.0)	18.0	(6.0)	21.3	(5.7)	26.2	(3.2)	28.5	(2.3)

Tabel 2. Beschrijvende statistiek voor de beoordelingsniveaus van de spontane taalproductie en de scores voor de opdrachtgroepen en subtests van de AAT. In de eerste regel staan steeds de mediaan en het bereik, in de tweede regel het gemiddelde en de standaardafwijking.

Resultaten

De beschrijvende statistiek voor de zes beoordelingsniveaus van de spontane taalproductie en voor de afzonderlijke subtests en opdrachtgroepen van de AAT is in tabel 2 voor elke proefpersoongroep weergegeven.

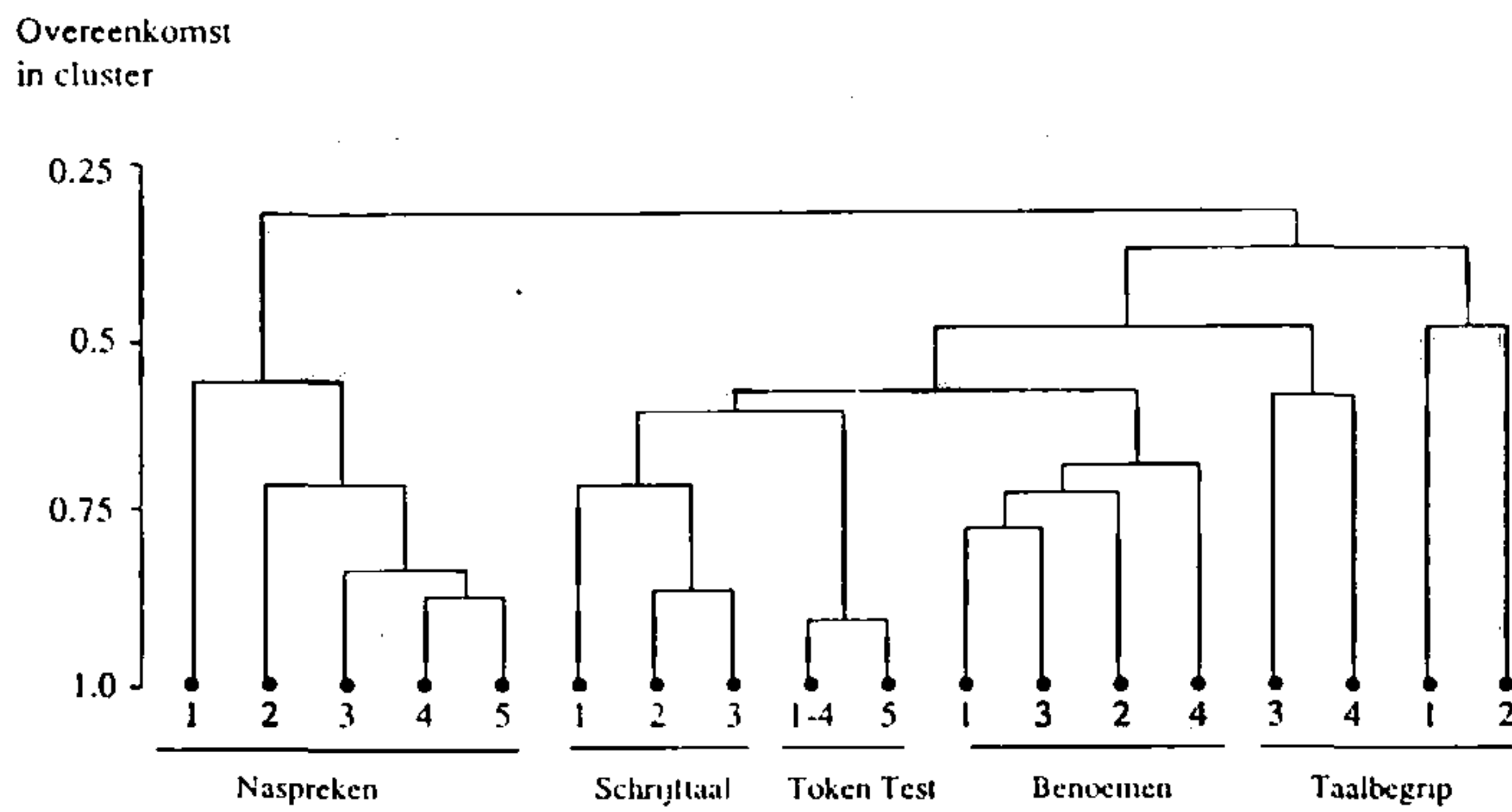
In het vervolg van deze paragraaf beperken wij ons tot de resultaten van de belangrijkste analyses met betrekking tot de constructvaliditeit, differentiële

validiteit en betrouwbaarheid.

Constructvaliditeit

De eerste vraag betreft de rechtvaardiging van de linguïstisch gemotiveerde combinatie van opdrachten in opdrachtgroepen en van deze opdrachtgroepen in subtests. Om te onderzoeken in hoeverre tussen opdrachtgroepen binnen eenzelfde subtest sterkere samenhangen bestaan dan tussen opdrachtgroepen van verschillende subtests, is de intercorre-

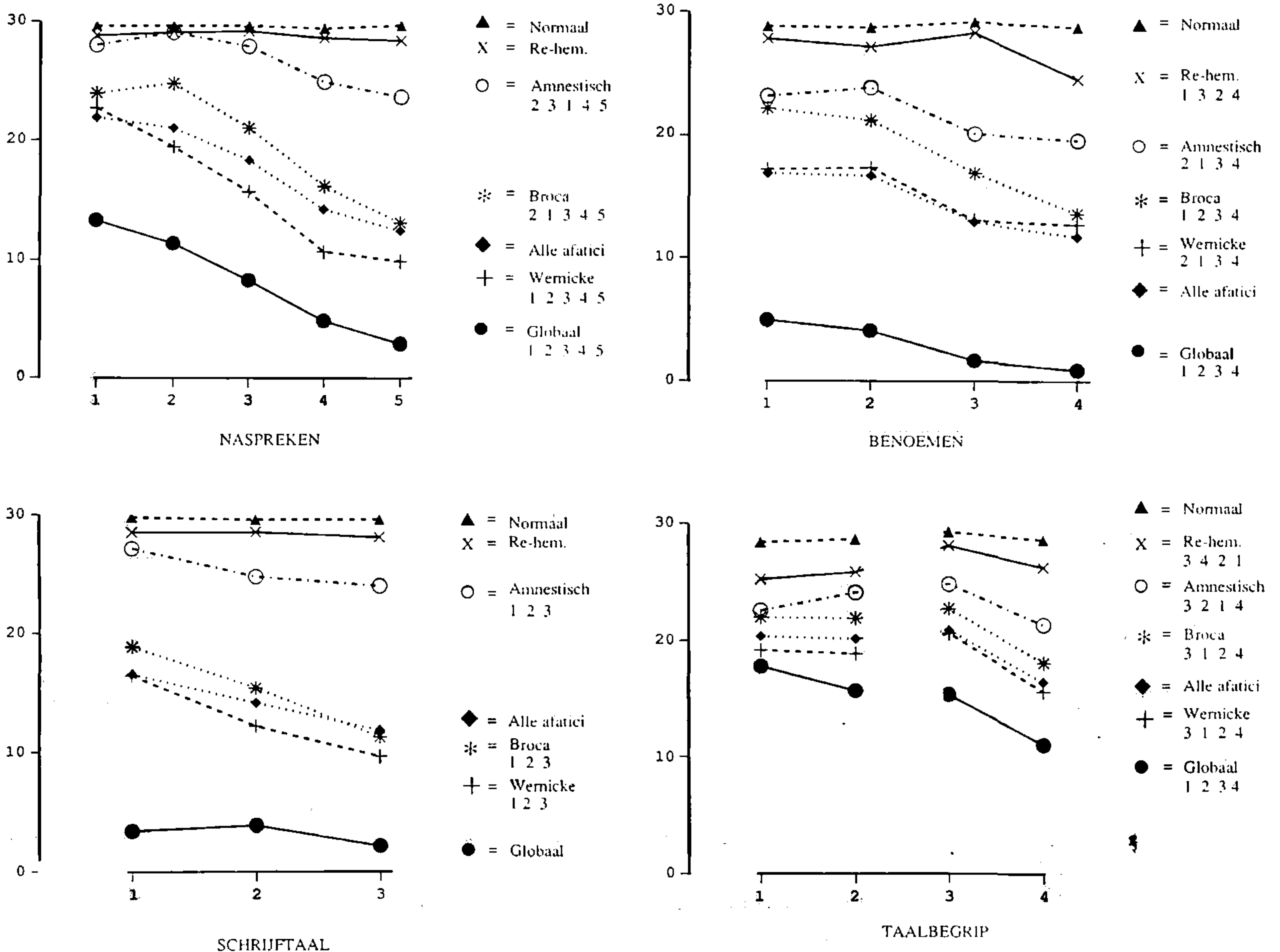
latiematrix van alle opdrachtgroepen aan een hiërarchische clusteranalyse onderworpen. Het resultaat van deze analyse voor alle 120 afasiepatiënten samen is in figuur 1 als boomdiagram weergegeven.



Figuur 1. Groepering van de opdrachtgroepen van de AAT (hiërarchische clusteranalyse met complete linkage procedure (Hartigan, 1975) voor μ^2 -coëfficiënten (Raveh, 1978)).

Met uitzondering van de subtest Taalbegrip clusteren de opdrachtgroepen van een subtest eerst met elkaar voordat ze met een andere subtest worden samengevoegd. Net als voor de Duitse AAT komt binnen de subtest Naspreken de eerste opdracht-

groep (klanken) het minst overeen met de overige opdrachtgroepen. Hiervan komen de groepen met de meest complexe stimuli, de delen 4 en 5 (samengestelde nomina en zinnen), het sterkst met elkaar overeen. Daaraan worden achtereenvolgens de groepen met afnemende complexiteit, de delen 3 en 2 (leen- en vreemde woorden, éénlettergrepige woorden), toegevoegd. Zoals verwacht vertonen binnen de subtest Schrijftaal de delen 2 en 3 (samenvoegen en schrijven op dictaat) sterkere samenhang met elkaar dan met het hardop lezen (deel 1). In de subtest Benoemen is het beschrijvend benoemen in deel 4 (zinnen) minder vergelijkbaar met de overige delen, die benoemen in een engere zin vereisen. Hier wijkt het benoemen van kleuren (deel 2) qua samenhang wederom licht af van het benoemen van lijntekeningen van objecten met enkelvoudige (deel 1) respectievelijk samengestelde nomina (deel 3). Als geheel vertonen de delen binnen de subtest Taalbegrip tenslotte de minst sterke overeenkomst met elkaar. De indeling naar auditief (deel 1 en 2) en leesinhoudelijk taalbegrip (deel 3 en 4) is echter duidelijk aanwezig. Het laatstgenoemde deelcluster wordt dan aan de drie subtestclusters toegevoegd, die het sterkst een beroep doen op centrale taalverwerkingsprocessen. Als laatste wordt het Naspreken met alle andere opdrachtgroepen verbonden.



Figuur 2. Gemiddelde beoordelingen van de spontane taalproductie en somwaarden voor de opdrachtgroepen en subtests van de AAT voor alle groepen.

Eén van de belangrijkste constructieprincipes van de AAT is de *toenemende linguïstische complexiteit* (a) tussen opdrachtgroepen binnen subtests, en (b) tussen opdrachten binnen opdrachtgroepen. In strikte zin geldt dit voor de Token Test, het Naspreken en het Benoemen. In de subtests Schrijftaal en Taalbegrip worden verschillende aspecten van deze taal-functies onderzocht. Hier zijn deze aspecten naar verwachte moeilijkheidsgraad gerangschikt.

Om het constructie-aspect (a) te toetsen zijn met behulp van verdelingsvrije permutatietoetsen (Edgington, 1987) de prestaties op de opdrachtgroepen per subtest met elkaar vergeleken. In tabel 2 en figuur 2 zijn de gemiddelden per opdrachtgroep en subtest voor de totale groep van 120 afasiepatiënten en voor alle afzonderlijke proefpersoongroepen weergegeven.

Ten eerste is voor elke subtest onderzocht, of de oplopende talige moeilijkheidsgraad in alle patiëntengroepen hetzelfde is. De permutatietoets (Pyhel, 1980; Willmes, 1982) laat significante interacties tussen proefpersoongroep en opdrachtgroep zien voor alle subtests. Daarom zijn vervolgens steeds eenvoudige hoofdeffecten getoetst met aansluitende multiële vergelijkingen (Petrondas en Gabriel, 1983). De resultaten zijn in tabel 3 weergegeven. Proefpersoongroepen en opdrachtgroepen, die niet significant van elkaar verschillen zijn door onderstreping met elkaar verbonden.

men twee voorbeelden. De bodemeffecten zijn voor patiënten met een globale afasie vooral voor de subtest Schrijftaal duidelijk te zien. In de subtest Taalbegrip zijn alleen voor het leesinhoudelijk begrijpen zinnen als geheel significant moeilijker te verwerken dan woorden. Voor het auditieve begrijpen bestaat dit verschil niet.

Het constructieprincipe (b) is de toenemende complexiteit van opdrachten binnen elke opdrachtgroep. Figuur 3 toont de moeilijkheid van opdrachten per opdrachtgroep voor alle afasiepatiënten.

Globaal gezien vertonen de curves een dalende tendens die het principe van systematisch stijgende moeilijkheidsgraad bevestigt. Per subtest zijn ook kleine afwijkingen van deze tendens zichtbaar. Een duidelijke afname qua prestatie is voor de opdrachten 6-10 in deel 3 van het Naspreken en voor de opdrachten 1-8 van de delen 4 en 5 te zien. De laatste twee opdrachten van de delen 2 en 4 schijnen daarentegen iets te gemakkelijk te zijn. In deel 3 van de subtest Schrijftaal neemt de moeilijkheid met uitzondering van de uitzonderlijk moeilijke opdracht 3 continu toe. Bij het hardop lezen en samenstellen op dictaat vindt na steeds vijf opdrachten een breuk plaats. Bij het samenstellen op dictaat wordt dan van letters en lettercombinaties naar woorddelen en woorden overgegaan. Het Benoemen van kleuren (deel 2) vertoont grotendeels de verwachte toename in moeilijkheid. In deel 1 is elke opdracht met een even nummer makkelijker dan zijn voorganger met een oneven nummer. Voor de delen 3 en 4 is geen duidelijke toename van de moeilijkheid van de opdrachten vast te stellen. In de beide delen van het auditieve Taalbegrip bestaan er deels ongewoon grote verschillen qua moeilijkheid tussen de opdrachten, tot verschillen van meer dan één puntwaarde toe.

Zoals verwacht bestaat er binnen de delen 1-4 van de Token Test (niet in figuur 3) geen grote systematische variabiliteit in de moeilijkheid van opdrachten. Alleen in deel 5 corresponderen de verschillen in syntactische complexiteit met verschillende prestaties. Evenals voor de Duitse versie van de Token Test (Willmes, 1981) zijn de opdrachten 7 en 10 moeilijk: hierin is de syntactische volgorde omgekeerd aan de beoogde volgorde van uitvoering. Bovendien is opdracht 2, waarin op de exclusieve interpretatie van het woord 'of' gedoeld wordt, extreem moeilijk.

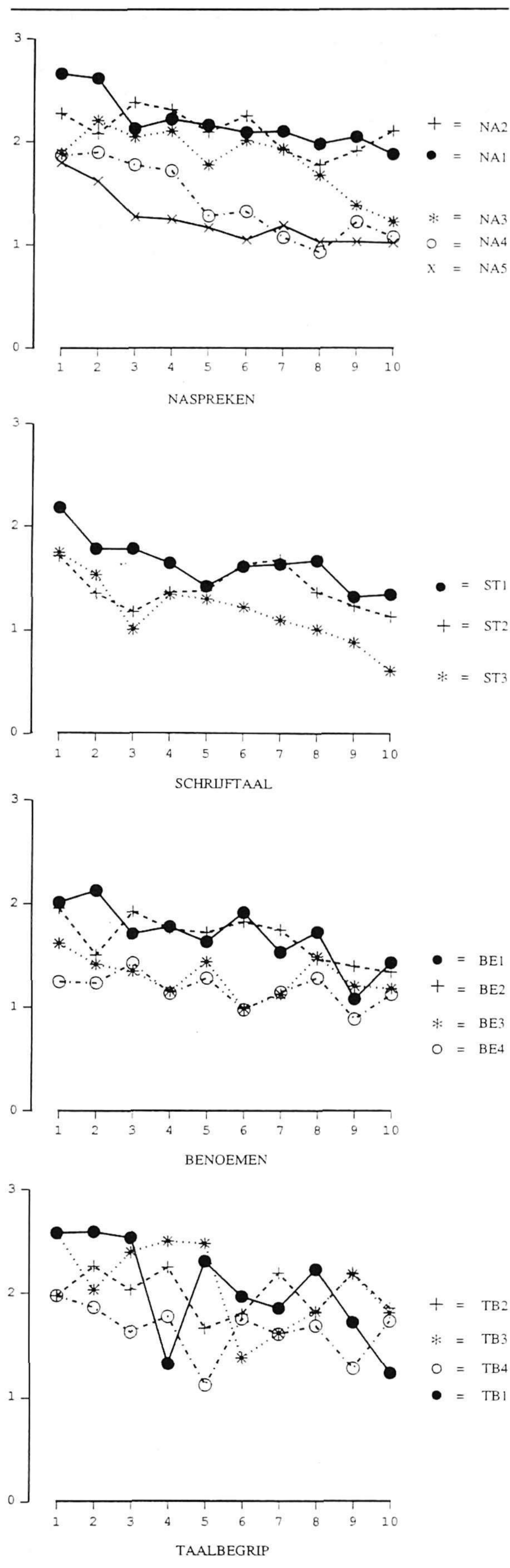
Differentiële validiteit

Als eerste stap zijn de verschillen tussen de groepen met betrekking tot de totale prestaties bepaald. Voor de beide controlegroepen, patiënten met rechter hersenhelftbeschadiging en proefpersonen zonder hersenbeschadiging, is geen analyse van de spontane taalproductie uitgevoerd. De schalen zijn voor de beoordeling van afatische stoornissen ontworpen, zodat dialectvarianten of vereenvoudigingen in de beoordeling buiten beschouwing blijven. Bij de subtests vertonen de afasiepatiënten een veel grotere systematische variabiliteit in prestaties dan de controlegroepen. De resultaten van de permutatietoetsen volgens Petrondas en Gabriel (1983) staan in

a)				
Pp-groep	Opdrachtgroep			
	Naspreken	Schrijftaal	Benoemen	Taalbegrip
Globaal	<u>1 2 3 4 5</u>	<u>2 1 3</u>	<u>1 2 3 4</u>	<u>1 2 3 4</u>
Wernicke	<u>1 2 3 4 5</u>	<u>1 2 3</u>	<u>2 1 3 4</u>	<u>3 1 2 4</u>
Broca	<u>2 1 3 4 5</u>	<u>1 2 3</u>	<u>1 2 3 4</u>	<u>3 1 2 4</u>
Amnestisch	<u>2 1 3 4 5</u>	<u>1 2 3</u>	<u>2 1 3 4</u>	<u>3 2 1 4</u>
Re-hemisfeer	<u>3 2 1 4 5</u>	<u>1 2 3</u>	<u>3 1 2 4</u>	<u>3 4 2 1</u>
Normaal	<u>1 2 3 5 4</u>	<u>1 2 3</u>	<u>3 2 1 4</u>	<u>3 4 2 1</u>
b)				
Opdrachtgroep	Pp-groep			
	Naspreken	Schrijftaal	Benoemen	Taalbegrip
1	<u>G W B A R N</u>	<u>G W B A R N</u>	<u>G W B A R N</u>	<u>G W B A R N</u>
2	<u>G W B R A N</u>	<u>G W B A R N</u>	<u>G W B A R N</u>	<u>G W B A R N</u>
3	<u>G W B A R N</u>	<u>G W B A R N</u>	<u>G W B A R N</u>	<u>G W B A R N</u>
4	<u>G W B A R N</u>		<u>G W B A R N</u>	<u>G W B A R N</u>
5	<u>G W B A R N</u>			<u>G W B A R N</u>

Tabel 3. Verschillen in prestaties a) tussen de opdrachtgroepen van de linguïstische subtests per proefpersoongroep (pp-groep) en b) respectievelijk tussen de pp-groepen per opdrachtgroep van de AAT. De opdrachtgroepen zijn geordend volgens opklimmende gemiddelde moeilijkheid, de pp-groepen volgens opklimmende prestatie; niet-significante verschillende groepen zijn door onderstreping met elkaar verbonden.

De resultaten laten zien, dat aan de genoemde verwachtingen goeddeels wordt voldaan. Tussen de opdrachtgroepen blijken de prestaties afhankelijk van talige moeilijkheidsgraad af te nemen. Anderzijds zijn er taalmodaliteiten waarin de prestaties van bepaalde groepen van afasiepatiënten ofwel voor alle opdrachtgroepen goed ofwel voor alle opdrachtgroepen slecht zijn, hoewel de opdrachten linguïstisch gezien qua moeilijkheid verschillen. We noe-



Figuur 3. Moeilijkheden van de items in de opdrachtgroepen van de subtests van de AAT alle afatici (n=120).

tabel 4 voor respectievelijk de spontane taalproductieschalen en de subtests. De door een streep met elkaar verbonden proefpersoongroepen zijn niet significant van elkaar te onderscheiden.

Spontane taalproductie				
COM	G	<u>B</u>	<u>W</u>	A
ART	G	B	<u>A</u>	<u>W</u>
AUT	G	<u>W</u>	<u>B</u>	A
SEM	G	<u>W</u>	<u>B</u>	A
FON	G	<u>B</u>	<u>W</u>	A
SYN	G	B	W	A

Subtests						
Token Test	<u>G</u>	<u>W</u>	<u>B</u>	A	<u>R</u>	<u>N</u>
Naspreken	G	<u>W</u>	<u>B</u>	A	R	N
Schrijftaal	G	<u>W</u>	<u>B</u>	A	R	N
Benoemen	G	<u>W</u>	<u>B</u>	A	R	N
Taalbegrip	<u>G</u>	<u>W</u>	<u>B</u>	A	R	N

Tabel 4. Verschillen in prestaties tussen de proefpersoongroepen voor de beoordelingsschalen van de spontane taalproductie en de subtests van de AAT. De proefpersoongroepen zijn geordend volgens opklimmende prestatie; niet-significant verschillende groepen zijn door onderstreping met elkaar verbonden.

Broca- en Wernicke-patiënten verschillen in geen enkele subtest in hun kwantitatieve prestaties significant van elkaar. Dat is terug te voeren op het feit dat kwalitatief verschillende symptomen als even zware afwijkingen van de doelvorm beoordeeld worden. De beide controlegroepen scoren op alle subtests significant beter dan de afasiepatiënten. De gezonde proefpersonen hebben op alle testonderdelen behalve de Token Test significant hogere scores dan de proefpersonen met rechter hersenhelftbeschadiging.

Voor de schalen van de *spontane taalproductie* is de beschrijvende statistiek reeds in tabel 2 gegeven. De gemiddelde prestaties vertonen per afatische groep eenzelfde patroon als bij de Duitse AAT. In het algemeen wijken de gemiddelden niet meer dan 0.5 van de gemiddelde Duitse beoordelingen af. Duidelijker verschillen bestaan er voor de patiënten met globale afasie, die met een gemiddelde van 0.2 voor de fonematische structuur veel sterker gestoord zijn dan de patiënten met globale afasie in de Duitse steekproef met een gemiddelde van 1.5 op deze schaal. Ook de Nederlandstalige patiënten met amnestische afasie zijn iets sterker gestoord voor de niveaus communicatief gedrag (3.3 tegenover 4.1), geautomatiseerde taal (3.9 tegenover 4.5) en syntactische structuur (3.4 tegenover 4.1). De syndroomverschillen per beoordelingsschaal van de spontane taalproductie zijn, naast de subtests, in tabel 4 weergegeven. Het patroon van verschillen komt sterk met dat van de Duitse versie overeen. Per schaal zijn verschillende groeperingen zichtbaar, die de specifieke eigenschappen van de stoornissen van de afatische syndromen weerspiegelen. Voor de schaal ter

beoordeling van de syntactische structuur verschillen alle afatische groepen significant van elkaar. Voor de bepaling van *het discriminerend vermogen* van de AAT is vervolgens onderzocht, hoe goed de externe syndroomclassificatie van de patiënten overeenstemt met statistische groeperingsprocedures van alle AAT-variabelen.

Aangezien de ruwe dataverdelingen in de groepen sterk van normale verdelingen afwijken, is een non-parametrische discriminantanalyse gebruikt (programma ALLOC80, Hermans e.a., 1982). Als eerste is getoetst, hoe goed de AAT de vraag 'afatisch ja/nee?' kan beantwoorden. Met andere woorden: kunnen de AAT-scores de externe scheiding tussen afasiepatiënten en controleproefpersonen (door ervaren beoordelaars) reproduceren? Dit reclassificatievermogen van de AAT is eerst voor alle subtests bepaald. Als criterium is hierbij, net als bij de Duitstalige AAT, een aposteriori-classificatiewaarschijnlijkheid van minstens 80% gehanteerd; bij waarschijnlijkheden tussen 50% en 80% heet de classificatie twijfelachtig. Daarna is met een zogenaamde stapsgewijze variabelenselectie onderzocht, welk deel van de subtests reeds een zo goed mogelijke scheiding van de groepen oplevert. De best discriminerende subtest blijkt de Token Test te zijn, op de voet gevolgd door de subtests Benoemen of Schrijftaal. De subtest die vervolgens in combinatie met de Token Test de grootste verbetering in de reclassificatie bewerkstelligt is de subtest Naspreken. Het aantal correct (dat wil zeggen in overeenstemming met de externe beoordeling) gereclassificeerde proefpersonen neemt met meer subtests niet verder toe. De resultaten staan in tabel 5.

ALLOC80* classificatie	Externe klinische diagnose		Reclassificatie
	afasie	geen afasie	
a) Token Test en Naspreken			
afasie	108	1	92.5%
afasie?***	3 (2B, 1A)	2	
geen afasie?	4 (4A)	3	95.0%
geen afasie	5 (2W, 3A)	54	
b) alle subtests			
afasie	108	2	94.2%
afasie?	5 (5A)	2	
geen afasie?	2 (1B, 1A)	3	93.3%
geen afasie	5 (2W, 3A)	53	

* Basisverhouding afasie – geen afasie 1/3:2/3

** Classificatiewaarschijnlijkheid tussen 50 en 80%

Tabel 5. Classificatie van afatische en niet-afatische proefpersonen aan de hand van een non-parametrische discriminant analyse (ALLOC80) voor de subtests van de AAT: a) alleen Token Test en Naspreken, b) alle subtests.

In de laatste kolom van deze tabel staan de percentages correcte reclassificatie volgens het zogenaamde 'mildere' classificatiecriterium; hierbij zijn de classificaties volgens het strenge criterium (waarschijnlijkheden van tenminste 80%) en de twijfelachtige classificaties (waarschijnlijkheden tussen 50% en 80%) samengenomen. In dat geval worden zowel voor de geselecteerde variabelen als voor alle vijf de subtests

steeds duidelijk meer dan 90% van de afasiepatiënten en niet-afasiepatiënten correct gereclassificeerd. Alle incorrect gereclassificeerde niet-afatische patiënten zijn patiënten met een rechter hersenhelftbeschadiging. De twee incorrect gereclassificeerde afasiepatiënten blijken Wernicke-patiënten te zijn, die ongewoon goede prestaties op de subtests vertonen. Hun spontane taalproductie is echter evident afatisch. Over het geheel genomen is hiermee voor de Nederlandstalige versie van de AAT een goed discriminerend vermogen aangetoond.

Vervolgens is getoetst, in hoeverre de externe classificatie naar afatische standaardsyndromen aan de hand van de AAT-prestaties reproduceerbaar is. Er zijn analyses voor de zes spontane taalproductieschalen, voor de vijf subtests en voor alle elf AAT-variabelen gezamenlijk uitgevoerd. Als classificatiecriterium is net als voor de Duitse versie een streng classificatiecriterium van tenminste 70% gehanteerd. De resultaten staan in tabel 6.

ALLOC80* classificatie	Externe klinische diagnose							
	G	W	B	A	G	W	B	A
a) Alleen spontane taalproductie	alle variabelen gereclassificeerd: 80.8% (76.7%)◆				geselecteerde variabelen: SYN, AUT gereclassificeerd: 88.3% (77.5%)◆			
G	28	1	–	–	29	1	–	–
G?*	G/B:1	–	–	–	G/B:1	–	–	–
W	1	19	–	6	–	23	–	6
W?	–	W/A:2	–	W/A:6	–	–	–	–
B	–	1	29	1	–	–	29	–
B?	–	–	–	–	–	–	–	–
A	–	5	1	16	–	–	1	12
A?	–	A/B:1	–	A/W:1	–	A/W:6	–	A/W:12
b) Spontane taalproductie en subtests	alle variabelen gereclassificeerd: 85.0% (80.0%)◆				geselecteerde variabelen: SYN, AUT, TB gereclassificeerd: 89.2% (72.5%)◆			
G	30	2	–	–	29	1	–	–
G?	–	–	–	–	G/W:1	–	–	–
W	–	18	–	2	–	17	–	4
W?	–	W/A:1	–	W/A:2	–	W/A:9	–	W/A:4
B	–	1	28	1	–	–	29	–
B?	–	B/W:1	B/W:1	B/A:1	–	–	–	–
A	–	6	–	20	–	2	1	12
A?	–	A/W:1	A/W:1	A/W:4	–	A/W:1	–	A/W:10

* Gelijke basisverhoudingen (variabele kernschatters)

** Classificatiewaarschijnlijkheid tussen 50% en 70%, volgens de op-een-na-hoogste waarschijnlijkheid

◆ Reclassificatieresultaat tussen haakjes: behalve de gevallen met een ?

Tabel 6. Classificatie van afatische patiënten naar afasiesyndromen aan de hand van een non-parametrische discriminant analyse (ALLOC80): a) voor de zes spontane taalproductievariabelen, b) voor spontane taalproductie en subtests samen.

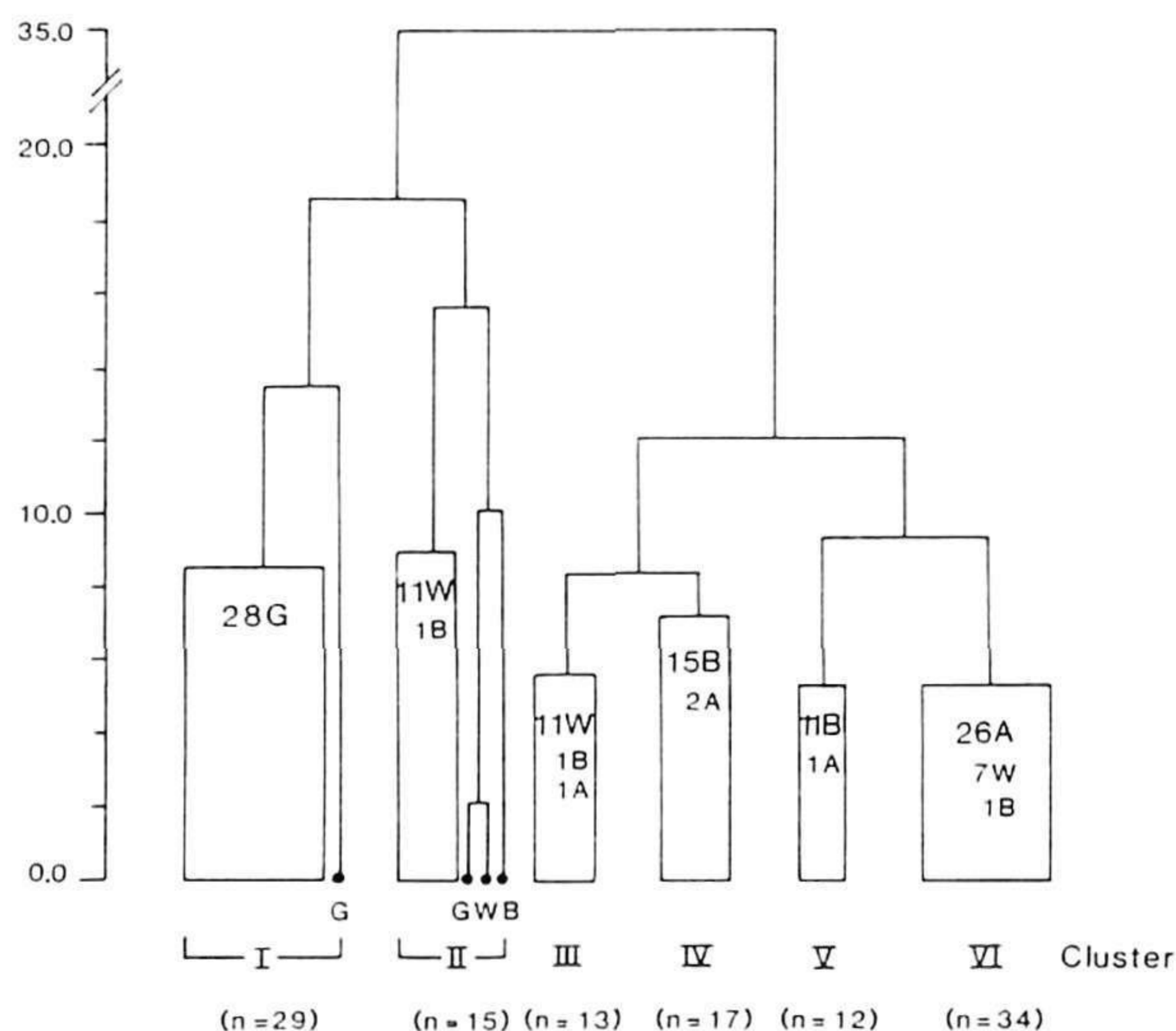
Voor de spontane taalproductie afzonderlijk (tabel 6a) worden, net als bij de Duitse AAT, de beoordelingsschalen syntactische structuur en geautomatiseerde taal geselecteerd. Toepassing van het mildere (respectievelijk strengere) classificatiecriterium levert een reclassificatieresultaat van 88.3% (77.5%) op. Wanneer alle variabelen van de spontane taalbeoordeling simultaan gebruikt worden, daalt dit resultaat naar 80.8% (76.7%). Opvallend is

de redelijk hoge overlapping tussen de patiënten met Wernicke- en amnestische afasie. De ene Wernicke-patiënt wiens spontane taalproductie als globaal-afatisch beoordeeld is, is een patiënt met overwegend neologistisch jargon.

De overeenkomst tussen klinische en ALLOC-classificatie voor de spontane taalproductie afzonderlijk is hoog. De vijf subtests daarentegen leveren een veel slechtere groepering op; de reclassificatie-overeenkomst bedraagt in dit geval slechts 60% (niet in tabel 6 opgenomen).

Bij beschouwing van spontane taalproductie en subtests samen (tabel 6b) wordt naast syntactische structuur en geautomatiseerde taal ook de subtest Taalbegrip geselecteerd. In totaal stijgt het reclassificatieresultaat slechts licht naar 89.2% (72.5%). Opvallend is het redelijk hoge aantal patiënten dat op grond van de spontane taalproductie door de beoordelaars als (zwaar) amnestisch gecategoriseerd wordt, terwijl deze patiënten hier als patiënten met Wernicke-afasie (deels met een zekere waarschijnlijkheid voor amnestische afasie) gereclassificeerd worden. Bij beschouwing van alle variabelen daalt het resultaat licht, naar 85% (80%).

Als volgende stap is onderzocht, in hoeverre de indeling naar de vier standaardsyndromen uit *alleen* de AAT-variabelen naar voren komt, dus zonder vergelijking met de externe classificaties. Het resultaat voor de spontane taalproductie en de subtests samen is in figuur 4 weergegeven. Ter vereenvoudiging zijn de eerste samenvattingen in de analyse niet meer expliciet afgebeeld.



Figuur 4. Groepering van de patiënten volgens overeenstemming van het AAT-profiel (hiërarchische clusteranalyse met average linkage procedure; gekwadrateerde euclidische afstanden van de z-getransformeerde waarden voor de spontane taalbeoordeling en de subtests).

Er blijken zes hoofdcusters te ontstaan. De groepen van mensen met globale of amnestische afasie nemen elk één cluster in beslag. Patiënten met de diagnose Wernicke- respectievelijk Broca-afasie verdelen zich over steeds twee clusters. Deze ver-

schillen voornamelijk in ernst van de stoornis van elkaar.

Betrouwbaarheid

Om te toetsen in hoeverre de afzonderlijke opdrachten van een subtest dezelfde (talige) prestatie vereisen, zijn consistentiecoëfficiënten (Cronbach's α -coëfficiënt, Cronbach, 1951) berekend. Ze zijn in tabel 7 voor de vijf subtests van de AAT weergegeven.

(n = 120)	(Cronbach's α)
Token Test	0.977
Naspreken	0.987
Schrijftaal	0.985
Benoemen	0.981
Taalbegrip	0.904
- Spreken	0.786 (0.880)*
- Lezen	0.879 (0.936)*

* Revaluatie volgens de formule van Spearman-Brown.

Tabel 7. Betrouwbaarheid (interne consistentie) van de AAT-subtests: Cronbach's α (Cronbach, 1951).

Alle coëfficiënten zijn hoger dan 0.90, de meeste liggen zelfs rond 0.98, behalve voor de subtest Taalbegrip. Voor deze subtest is daarom gekeken naar een mogelijke discrepantie tussen auditief en leesinhoudelijk taalbegrip. De betrouwbaarheidsschatting voor het leesinhoudelijk begrip afzonderlijk is bevredigend, alleen het auditief begrip is minder consistent. Ook bij de analyses met betrekking tot de constructvaliditeit kwam het auditieve deel van de subtest Taalbegrip steeds als minst homogeen testonderdeel uit de bus. De mogelijke redenen komen in de discussie aan de orde.

Discussie

Constructvaliditeit

Zoals door Graetz e.a. (1991) is beschreven, zijn de opdrachten van de AAT volgens specifieke linguïstische criteria geconstrueerd. Vanwege de hoge graad van verwantschap tussen de Duitse en de Nederlandse taal is de adaptatie van de opdrachten goed mogelijk gebleken, zonder dat de algemene linguïstische parameters van een opdracht veranderd moesten worden. Dit was voor de Italiaanse versie gedeeltelijk wel noodzakelijk.

De subtests zijn naar taalmodaliteiten opgebouwd en bevatten steeds verschillende linguïstische eenheden, die op hun beurt naar toenemende linguïstische complexiteit zijn gerangschikt. De analyses met betrekking tot de constructvaliditeit van de Nederlandstalige versie van de AAT hebben laten zien, dat een groot deel van de constructieprincipes empirisch aantoonbaar is. De boomdiagram in figuur 1 laat zien, dat de combinatie van opdrachten in groepen enerzijds en van opdrachtgroepen in subtests anderzijds gerechtvaardigd is. Voor de beide subtests Naspreken en Schrijftaal, waarin de toenemende complexiteit het voornaamste constructieprincipe is,

blijkt uit de geobserveerde moeilijkheden van de opdrachten, dat complexiteit van consonantclusters, toename in aantal lettergrepen alsmede toename in aantal morfemen, respectievelijk woorden, inderdaad een effect hebben binnen de beide subtests als geheel. Ook voor de afzonderlijke opdrachtgroepen van deze twee subtests Naspreken en Schrijftaal is het principe van de toenemende complexiteit verregaand bevestigd. Het effect van de complexiteit is in de opdrachtgroepen van de subtest Benoemen alleen voor het benoemen van kleuren duidelijk. Binnen de subtest Taalbegrip tenslotte geldt voor alle afatische groepen alleen voor het leesinhoudelijk taalbegrip, dat woorden makkelijker te verwerken zijn dan zinnen. Er kon niet worden bevestigd dat het begrijpen van woorden en zinnen makkelijker is wanneer zij auditief worden aangeboden, dan wanneer zij schriftelijk worden aangeboden. Op de enigszins afwijkende positie van deze subtest gaan we nog in.

Differentiële validiteit

De eerste eis die aan een afasietest gesteld wordt, is het kunnen onderscheiden van afasiepatiënten van niet-afatische controle-patiënten. Aan deze eis wordt door de subtests van de AAT in ruim 93% van de gevallen voldaan. De subtest die numeriek gezien de beste scheiding maakt, is de Token Test. Samen met het Naspreken vertoont de Token Test de beste selectieve eigenschappen. De analyses hebben aangetoond, dat er vooral voor de zes spontane taalproductievariabelen en gedeeltelijk ook voor de subtests meerdere constellaties van verschillen tussen de afatische groepen bestaan. De controlegroepen zijn in alle subtests significant beter dan elk van de afatische groepen. Volgens het mildere classificatiecriterium vertonen behalve voor de subtest Schrijftaal alle andere subtests een scoreprofiel met significante verschillen tussen alle afatische groepen. Ook volgens het strengere classificatiecriterium vertoont de groep patiënten met globale afasie overal de significant slechtste prestaties. Voor de drie groepen van patiënten met Wernicke-, Broca- of amnestische afasie zijn de relaties tussen de prestaties onderling niet voor alle variabelen gelijk. Weliswaar geldt dat de patiënten met Wernicke-afasie meestal numeriek de slechtste prestaties en de patiënten met een amnestische afasie de beste prestaties leveren. Bij analyse van deze verschillen valt het volgende op: de patiënten met een amnestische afasie zijn in de algemene inschatting van het communicatief gedrag alsmede op de niveaus van geautomatiseerde taal, fonematische en syntactische structuur significant beter dan alle andere afasiepatiënten. De patiënten met een afasie van Broca halen als groep alleen voor de beide spontane taalproductieniveaus articulatie en prosodie, respectievelijk syntactische structuur, gemiddeld lagere scores dan de patiënten met Wernicke-afasie. Terwijl het eerste verschil tot de vaak optredende spraakmotorische stoornissen bij Broca-afasie kan worden herleid, dient men voor het beoordelingsniveau van de syntactische structuur het volgende te bedenken. Dat de syntactische struc-

tuur bij patiënten met Wernicke-afasie 'beter' bewaard gebleven lijkt dan bij Broca-patiënten, is mede een consequentie van de constructie van deze beoordelingsschaal. Op deze schaal is het ontbreken van complexe zinsbouw alsmede van functiewoorden en flexievormen aanvankelijk als ernstiger afwijkend beschouwd dan complexe foutieve zinsbouw en foutieve toepassing van functiewoorden en flexievormen. De laatste jaren staat de (on-)grammaticaliteit van de uitingen van Broca-patiënten echter ter discussie (bijvoorbeeld Kolk en Hofstede, 1988).

Belangrijker dan het aantonen van gemiddelde verschillen is het aantonen dat objectieve groeperingsverschillen goed met de klinische classificatie van patiënten overeenkomen. In totaal levert de gehele AAT (elf variabelen) een goede overeenstemming op van 80% bij de patiënten met standaard afasiesyndromen. Dit percentage wekt beslist geen te hoge verwachtingen met betrekking tot het classificatievermogen van de AAT. Ten eerste zijn alleen gevallen met een hoge classificatiewaarschijnlijkheid (hoger dan 70%) als correct gereclassificeerd beschouwd. In de tweede plaats wordt de hoeveelheid incorrecte classificatie laag gehouden, doordat de ALLOC-procedure de zogenaamde 'leaving-one-out' methode hanteert (Lachenbruch, 1977). Dit betekent dat bij de bepaling van de discriminantfunctie beurtelings elke te classificeren patiënt zelf niet in de steekproef wordt opgenomen. Zodoende kunnen diens specifieke eigenschappen de groepering niet beïnvloeden. Bovendien is het grootste deel van de niet-overeenstemmende gevallen afasiologisch goed verklaarbaar.

De meest aan de classificatie bijdragende variabelen blijken de spontane taalproductieniveaus ter beoordeling van de syntactische structuur en de geautomatiseerde taal te zijn. Daarom is het niet verwonderlijk dat de zes spontane taalproductievariabelen alleen een bijna even groot classificatievermogen bezitten als alle AAT-variabelen samen. Overigens geldt voor zowel het algemene discriminerende vermogen van de AAT als voor de syndroomclassificaties dat de gevonden resultaten zeer sterk met die bij de Duitse en Italiaanse versies van de AAT overeenkomen. Bij de Nederlandstalige AAT lukt classificatie van de patiënten met een Broca-afasie zelfs iets beter dan in de Duitse versie. Daarentegen lukt in de Duitse versie de scheiding tussen patiënten met Wernicke- of amnestische afasie beter.

Betrouwbaarheid

De hoogte van de consistentiecoëfficiënten laat zien, dat de subtests de talige prestaties zeer betrouwbaar meten. Bovendien komen de gevonden coëfficiënten ook numeriek sterk met die van de Duitse en Italiaanse AAT-versies overeen. Iets minder goed, maar altijd nog hoger dan 0.90 is de coëfficiënt voor de gehele subtest Taalbegrip. Men dient zich derhalve af te vragen, waar het aan ligt, dat bij deze en enkele andere analyses de subtest Taalbegrip en daarvan vooral het auditieve deel minder homogeen is. In de AAT zijn bewust geen opdrachten opgeno-

men, die motorisch complexere handelingen vereisen, vanwege een mogelijke interactie met stoornissen van de planning en/of de uitvoering van een handeling. Daarom is bij de subtest Taalbegrip met het oog op de antwoordmodus gekozen voor meerkeuze-opdrachten. Vanwege de metalinguïstische aspecten van deze opdrachten is in vergelijking met de andere modaliteiten niet uit te sluiten dat voor de uitvoering geheel andersoortige, ook niet-talige strategieën gebruikt kunnen worden. De opdrachten vereisen bovendien complexe semantische en cognitieve verwerking. Tenslotte kan het gevonden resultaat echter ook evidentie zijn voor de vaak beweerde bredere representatie van lexicaal-semantische informatie in de hersenen (Chiarello, 1988). Voor de verdere afwijking van het auditieve begrip ten opzichte van het leesinhoudelijke begrip kan de andere presentatievorm van de stimuli verantwoordelijk zijn. Ter verwerking van de geschreven stimulus in het leesinhoudelijk deel staat de proefpersoon meer tijd ter beschikking dan ter verwerking van de gesproken stimulus in het auditieve deel. In het leesinhoudelijke deel is de schriftelijke stimulus als geheel langer aanwezig. Ook kan de proefpersoon zelf bepalen, wanneer de meerkeuzebladzijde wordt voorgelegd. Dit maakt het mogelijk om de stimulus bijvoorbeeld in een andere volgorde ter verwerking op te nemen, terwijl bij het auditieve deel de stimulus alleen in de aangeboden volgorde kan worden opgenomen. Het inconsistente antwoordpatroon bij het auditieve deel is daarom wellicht op het onvolledig ter beschikking hebben van informatie terug te voeren.

Vergelijking van de Nederlandse en de Duitse versie van de AAT

Eén van de doelstellingen bij de ontwikkeling van de Nederlandstalige versie van de AAT is geweest om betere gegevens voor contrastief onderzoek bij afasiepatiënten te verkrijgen. De zeer hoge overeenstemming tussen de beide testversies met betrekking tot de verschillende getoetste validiteits- en betrouwbaarheidsaspecten maakt dit met name voor contrastief-linguïstisch onderzoek van groepen van patiënten met vergelijkbare ernst van de stoornis en/of hetzelfde syndroom goed mogelijk. Over de toepasbaarheid van de testversies bij contrastief-linguïstische onderzoekingen bij individuele patiënten kan pas worden beslist, wanneer de Nederlandse normeringssteekproef tot hoge overeenkomsten blijkt te leiden. Daarvan wordt in het derde en laatste artikel verslag gedaan (Willmes e.a., in voorbereiding).

Conclusie

Twee eigenschappen van de AAT zijn niet voor de Nederlandstalige versie ervan opnieuw onderzocht. De scoringsobjectiviteit enerzijds en de test-hertestbetrouwbaarheid anderzijds waren reeds voor de Duitse versie zeer bevredigend aangetoond en er is geen reden om op deze gebieden verschillen in Nederland te verwachten. De objectiviteit van de afname wordt gewaarborgd door gelijke richtlijnen voor alle AAT-versies.

De overige eigenschappen, namelijk constructvaliditeit, differentiële validiteit en betrouwbaarheid in de zin van interne consistentie van de subtests, zijn wel aan de hand van een nieuwe, Nederlandse steekproef van 120 afasiepatiënten en 60 controleproefpersonen onderzocht en aangetoond.

De Nederlandstalige AAT is valide. De linguïstische constructieprincipes blijken zich op dusdanige wijze in de scores te weerspiegelen dat enerzijds de combinatie van opdrachten in groepen en van opdrachtgroepen in subtests gerechtvaardigd is. Anderzijds is ook het principe van de toenemende moeilijkheidsgraad van opdrachten binnen een groep alsmede van opdrachtgroepen binnen een subtest empirisch aangetoond. Ook blijkt de Nederlandstalige AAT goed in staat tot het maken van onderscheid tussen afatische patiënten en niet-afatische patiënten enerzijds en tussen de verschillende hoofdsyndromen anderzijds.

De Nederlandstalige AAT is betrouwbaar. De opdrachten en subtests zijn intern consistent, dat wil zeggen dat de eisen die de opdrachten aan de proefpersonen stellen van patiënt tot patiënt steeds dezelfde zijn. Over het geheel genomen is de consistentie van de subtests zo hoog, dat een betrouwbare bepaling van de ernst van de afatische stoornissen voor de individuele patiënt heel goed mogelijk is. Dit is op zijn beurt een eis waaraan voldaan moet zijn, wil men met methodes van psychometrisch verantwoorde individuele diagnostiek (Willmes, 1985) het AAT-profiel van een patiënt gedetailleerd met het oog op prestatieverschillen kunnen analyseren. Deze prestatieverschillen kunnen zowel verschillen tussen subtests binnen één AAT-afname betreffen alsook tussen twee AAT-profielen van eenzelfde patiënt. Analyse van deze beide mogelijkheden met behulp van een eenvoudig PC-programma wordt in het derde artikel uitvoerig toegelicht.

Op basis van de grotere heterogeniteit tussen de opdrachtgroepen met betrekking tot auditief en leesinhoudelijk taalbegrip is het overigens zinvol, om voor de individuele patiënt ook deze subtestdelen in het prestatieprofiel afzonderlijk te bekijken. Vanwege mogelijke dissociaties tussen het lees- en het schrijfvermogen bij afasiepatiënten is bovendien aan te raden om de delen van de subtest Schrijftaal eveneens afzonderlijk te beschouwen. Dit is toelaatbaar vanwege de hoge consistentie van de afzonderlijke opdrachtgroepen van zowel de subtest Schrijftaal als van het leesinhoudelijk Taalbegrip, die in het betrouwbaarheidsonderzoek is aangetoond. De uitspraken over het auditieve taalbegrip zullen met procedures van psychometrisch verantwoorde individuele diagnostiek echter minder exact moeten blijven. Ook de afzonderlijke analyses van de genoemde delen van de subtests Taalbegrip en Schrijftaal zijn een mogelijke optie binnen het genoemde PC-programma en komen in het laatste artikel uitgebreider aan de orde.

Aan het begin van de paragraaf 'Resultaten' is erop gewezen dat alleen de belangrijkste analyses met betrekking tot de constructvaliditeit, differentiële validiteit en betrouwbaarheid van de AAT behandeld worden. Gedetailleerde analyses met betrekking tot de constructvaliditeit met behulp van nonmetrische multidimensionale schaalprocedures ('Smallest Space Analyses-I', Lingoës, 1973) alsmede met betrekking tot de consistentie van de subtests met behulp van generaliseerbaarheidstoetsen (Cronbach e.a., 1972, programma GENOVA, Crick en Brennan, 1983) zijn met het oog op de lengte van dit artikel buiten beschouwing gelaten. Geïnteresseerde lezers worden uitgenodigd de resultaten van deze analyses aan te vragen bij P. Graetz.

Summary

In Part I (Graetz e.a., 1991) the linguistic properties of the original German version of the AAT have been described together with the criteria employed for the construction of the Dutch version. Different aspects of the validity and reliability of the Dutch AAT were examined in a sample of 120 left brain damaged patients and 60 controls. Construct validity could be demonstrated with respect to increased complexity of items within the different parts of subtests and similarity among parts within subtests. The Dutch AAT possesses high discriminatory power both between aphasic and non-aphasic subjects as well as between standard aphasic syndromes. Furthermore, the high internal consistency of the subtests allows a psychometrically sound assessment of single patients.

Literatuur

CHIARELLO, C., (1988). Semantic priming in the intact brain: Separate roles for the right and left hemispheres? In C. Chiarello (ed.). *Right hemisphere contributions to lexical semantics*. Berlin: Springer.

CRICK, J.E., R.L. BRENNAN, (1983). Manual for GENOVA: GENERALIZED Analysis Of VARIance System. *ACT Technical Bulletin 43*, Iowa City.

CRONBACH, L.J., (1951). Coefficient Alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

GRAETZ, P., R. DE BLESER, K. WILLMES, C. HEESCHEN, (1991). De Akense Afasie Test: Constructie van de Nederlandstalige versie van de AAT. *Tijdschrift voor Logopedie en Foniatrie*, 63, 3, 58-65.

GRAETZ, P., R. DE BLESER, K. WILLMES, (in voorbereiding). *Akense Afasie Test*.

HARTIGAN, J.A., (1975). *Clustering Algorithms*. New York: Wiley.

HERMANS, J., J.D.F. HABBEMA, T.K.D. KASANMOENTALIB, J.W. RAATGEVER, (1982). Manual for the ALLOC80 Discriminant Analysis Program. Leiden: Dept. of Medical Statistics, University of Leiden.

HUBER, W., K. POECK, D. WENIGER, K. WILLMES, (1983). *Der Aachener Aphasie Test*. Göttingen: Hogrefe.

KOLK, H.H.J., B.T.M. HOFSTEDE, (1988). Afasie en adaptatie. Een nieuwe kijk op afasiesymptomen.

Tijdschrift voor Revalidatiewetenschappen, 1, 4, 100-104.

KRIPPENDORFF, K., (1970). Estimating the reliability, systematic error, and random error of interval data. *Educational and Psychological Measurement*, 30, 61-70.

LACHENBRUCH, P.A., (1977). Some misuses of discriminant analysis. *Methods of Information in Medicine*, 4, 255-258.

LINGOËS, J.C., (1973). *The Guttman-Lingoës Non-metric Program Series*. Ann Arbor, Michigan: Mathesis Press.

ORGASS, B., (1986) *Der Token Test*. Weinheim: Beltz.

PETRONDAS, D.A., K.R. GABRIEL, (1983). Multiple comparisons by rerandomization tests. *Journal of the American Statistical Association*, 78, 949-957.

POECK, K., M. KERSCHENSTEINER, F.-J. STACHOWIAK, W. HUBER, (1975). Die Aphasien. *Aktuelle Neurologie*, 2, 159-169

PYHEL, N., (1980). Distribution-free r-sample tests for the hypothesis of parallelism of response profiles. *Biometrical Journal*, 22, 703-714.

RAVEH, A., (1978). Guttman's regression-free coefficients of monotonicity and polytonicity. In: S. Shye (ed.). *Theory Construction and Data Analysis in the Behavioral Sciences*, pp. 378-384, San Francisco: Jossey-Bass.

WENIGER, D., K. WILLMES, W. HUBER, K. POECK, (1981). Der Aachener Aphasie Test. Reliabilität und Auswertungsobjektivität. *Nervenarzt*, 52, 269-277.

WILLMES, K., (1981). A new look at the Token Test using probabilistic test models. *Neuropsychologia*, 19, 631-645.

WILLMES, K., (1982). A comparison between the Lehman & Wall rank tests and Pyhel's permutation test for the analysis of r independent samples of response curves. *Biometrical Journal*, 24, 714-722.

WILLMES, K., (1985). An approach to analyzing a single subject's scores obtained in a standardized test with application to the Aachen Aphasia Test (AAT). *Journal of Clinical and Experimental Neuropsychology*, 7, 331-352.

WILLMES, K., P. GRAETZ, R. DE BLESER, B. SCHULTE, A. KEYSER, (in voorbereiding). De Akense Afasie Test (AAT). III: Procedure voor individuele diagnostiek.

Ria De Bleser, linguïste en Klaus Willmes, psycholoog en mathematicus
Rheinisch-Westfälische Technische Hochschule
Neurologische Klinik, Pauwelsstraße, D - 5100
Aachen.

Patty Graetz, linguïste en Peter Hagoort, psycholoog
Max-Planck-Instituut voor Psycholinguïstiek
Postbus 310, 6500 AH Nijmegen.