

### Equivalentie en evaluatie van een conventionele en een computer-afname van een klachtenlijst

Theo K. Bouman en Wilmar B. Schaufeli\*

*In a group of 53 students a counterbalanced design was used to establish equivalence between paper-and-pencil and computer administrations of the Hopkins Symptom Check List. No differences in means and other distribution parameters were found between the two methods, nor did the correlation pattern between scales within each method diverge. The Semantic Differential showed significant higher scores on its three dimensions 'Evaluation', 'Potency' and 'Activity' for the concept 'Computer' as compared to the concept 'Testform'. In the discussion several remarks were made on the present study. Finally, standardization and quality control were recommended as prerequisites for the implementation of computerized psychological tests.*

#### INLEIDING

Evenals in veel andere gebieden van de samenleving heeft de (micro-)computer ook zijn intrede gedaan in de klinisch-psychologische diagnostiek. Het lijkt realistisch te verwachten dat de computer in de eerste plaats gebruikt zal worden als een machine waarmee de bestaande test-routines geautomatiseerd kunnen worden; pas daarna valt te denken aan innovatieve toepassingen zoals adaptief testen en het gebruik van *expert systems* (Butcher, Keller & Bacon, 1985). De voordelen van test-afnames per computer lijken evident: een snelle en foutloze scoring kan op gestandaardiseerde wijze plaatsvinden. Helaas is er in Nederland weinig programmatuur beschikbaar om dit te realiseren. Van Vliet en Bonarius concluderen in 1984 naar aanleiding van een enquête onder 244 psychologen naar het gebruik van de computer in de psychodiagnostiek: 'Er ontbreekt iedere vorm van op efficiëntie en produktiviteit gericht beleid in het ontwikkelen van software. De grote diversiteit in hardware, besturingssystemen en programmeertalen versterkt nogmaals de indruk van hobbyïsme en amateurisme' (p. 117-118). Recentelijk zijn er veranderingen te bespeuren door het beschikbaar komen van schil- en raamprogramma's zoals CDL (Computer Diagnostiek Leiden; Rombouts & Diekstra, 1987), GAME (Groningen Advanced Measurement Environment; Brokken, 1986), TL (Test Language; Reinders & Van Lenthe, 1986) en TAIGA (Twente Advanced Interactive Graphic Authoring system, ontwikkeld aan de Technische Universiteit Twente). Voorts vindt op enkele plaatsen in Nederland de psychodiagnostiek routinematig per computer plaats (bijvoorbeeld bij de Rijks Psychologische Dienst (Zaal, 1987) en in

de Sint Bavo in Noordwijkerhout (Wagenborg en Koning, 1984).

Wanneer de onderzochte persoon achter een beeldscherm en toetsenbord gezeten een test of vragenlijst invult, doet zich de vraag voor in hoeverre de resultaten vergelijkbaar c.q. equivalent zijn met de conventionele 'paper-and-pencil'-afname. Daarbij is het in tweede instantie van belang te weten, hoe respondenten de verschillende afnamesituaties beleven, teneinde mogelijke methodegebonden *respons-bias* (als een bedreiging voor de equivalentie) op het spoor te komen. Onze studie is een eerste aanzet tot de beantwoording van deze vragen.

#### EQUIVALENTIE VAN TESTSCORES

Men kan er niet op voorhand van uitgaan dat de resultaten van een gecomputeerde en een paper-and-pencil-afname van dezelfde test inwisselbaar, c.q. equivalent zijn. Om deze reden stelt de American Psychological Association (1986) in haar *Guidelines on Computer-Based Tests and Assessment*, dat pas bij gebleken equivalentie van testcores de normen van de conventionele versie gebruikt mogen worden voor de computer-uitvoering. Volgens deze richtlijnen zijn testcores equivalent wanneer aan de volgende twee eisen is voldaan:

- De rangordes van de scores van individuen die via alternatieve methoden zijn getest, dienen nauw overeen te komen.
- De gemiddelden en verdelingseigenschappen van scores dienen ongeveer gelijk te zijn, of door transformaties gelijk gemaakt te kunnen worden (APA, 1965, p. 18). In operationele termen betekent dit dat de twee versies zowel substantieel met elkaar dienen te correleren, alsook sterk op elkaar moeten lijken ten aanzien van gemiddelden, standaarddeviaties, kurtosis en scheefheid. Indien de equivalentie tussen de twee versies niet kan worden aangetoond, zal een hernormering van de computer-versie het noodzakelijke gevolg moeten zijn.

#### EQUIVALENTIE-ONDERZOEK

In Nederland is tot op heden nauwelijks equivalentie-onderzoek verricht. Publikaties behelzen voornamelijk beschouwingen over computertoepassingen, zoals een overzichtnummer van 'De Psycholoog' (1987, XXII) laat zien. Het tot dusverre – vooral in de Verenigde Staten – uitgevoerde equivalentie-onderzoek laat in de regel zien dat er geen of weinig verschillen bestaan tussen conventionele en computer-afnames van persoonlijkheidstests. Harrell & Lombardo (1984) namen bij 80 studenten de 16 Personality Factors (16PF) af in een *counterbalanced repeated measures design* en vonden dat de twee afnamemethoden equivalent waren. White, Clements & Fowler (1985) vermelden bij MMPi-afnames bij 150 studenten eveneens equivalentie schaalparameters. Katz & Dalby (1981) vonden bij 18 psychiatrische patiënten dat de schalen van de Eysenck Perso-

\* Vakgroep Klinische Psychologie, Rijksuniversiteit Groningen, Academisch Ziekenhuis, Oostersingel 59, 9713 EZ Groningen  
Met dank aan Laura van Bergen voor haar medewerking bij de afname en verwerking van de gegevens

nality Inventory geen verschil tussen afnamemethoden vertoonden. Een gecomputeriseerde versie van de Amsterdamse Biografische Vragenlijst (ABV, Wilde, 1970) leverde gemiddelde scores op bij 50 sollicitanten bij de Nederlandse Spoorwegen die vrijwel identiek waren aan de lokale normgroep (Knol & Schoonman, 1986). In dit laatste onderzoek moet equivalentie echter bij implicatie verondersteld worden, aangezien de proefpersonen slechts de gecomputeriseerde versie aangeboden kregen. Daarentegen vermelden Lushene, O'Neil & Dunn (1974) significante verschillen in gemiddelden op zeven van de dertien MMPI-schalen tussen een computer- en een boekjesvorm van de MMPI bij 63 studenten: de mediane test-herstest correlatie bedroeg .83 met een range van .47 tot .91. Ook Hermans, Hermans-Jansen & Van Gilst (1985) vonden bij 5 personen een verschil in respons op een standaard afgenomen en een computerversie van de Zelf Konfrontatie Methode (ZKM). De actieve rol die de testleider speelt bij de gangbare afname van de ZKM, is ongetwijfeld een belangrijke factor, die echter afwezig is bij een computer-afname. Van de Vijver (1987) laat in een overzicht zien dat er van equivalentie bij *speed*- en intelligentie-tests meestal geen sprake is. In het huidige onderzoek beperken we ons echter tot zelfbeoordelvragenlijsten.

#### KENMERKEN VAN PROEFPERSONEN

Een positieve waardering van de respondenten jegens de computer – al dan niet in vergelijking met een conventionele interview- of vragenlijstafname – vinden we eenstemmig in meerdere onderzoeken terug (Johnson & Williams, 1975; Angle et al., 1977; White, Clemens & Fowler, 1985; Harrell & Lombardo, 1984). Mathisen, Evans, Meyers & Kogan (1985) vonden geen verband tussen de subjectieve waardering van de computer-afname enerzijds en het bedieningsgemak anderzijds. De proefpersonen van White et al. (1985), evenals die van Harrell & Lombardo (1984) prefererden de afname per computer boven de schriftelijke variant. Lushene et al. (1974) maten bij hun proefpersonen echter een hoger niveau van toestandsangst in de conditie waarbij de computer als eerste afnamemethode werd gebruikt. Van belang is verder de bevinding van Cruickshank (1982) dat het stress-niveau van patiënten die *face-to-face* door een huisarts werden ondervraagd, met bijna 9% steeg ten opzichte van het niveau vóór het interview, tegenover een stijging van maar liefst 22% in het geval er van een computer gebruik werd gemaakt.

Bij het vergelijken van *face-to-face* en gecomputeriseerde interviews komen verschillende bevindingen naar voren ten aanzien van de openheid waarmee proefpersonen informatie verstrekken. Skinner & Allen (1983) onderzochten de equivalentie van een gecomputeriseerde en een *face-to-face* standaard intake-procedure in een verslavingssetting. De kwaliteit van de informatie bleek in beide gevallen identiek, de patiënten zijn bijvoorbeeld niet 'eerlijker' tegen de computer op vragen als 'Hoeveel glazen

drinkt u per dag?' dan tegen de menselijke interviewer. Wel verschilt de waardering van de patiënt voor de beide afnamen – gemeten met een semantische differentiaal –, waarbij de computer er opvallend positief van af komt; de machine wordt als meer 'relaxing' en 'interesting' beoordeeld dan de menselijke diagnosticus. Skinner & Allen (1983) wijzen verder op de rol van het opleidingsniveau van de respondenten; diegenen met een hogere opleiding waarden de gecomputeriseerde afname van het diagnostisch interview hoger dan de *face-to-face*-afname. Andere onderzoekers melden dat hun patiënten opener en minder defensief reageerden op de vragen die via de computer aangeboden werden, dan op vragen die door de clinicus gesteld werden (onder andere Greist et al., 1973).

In het algemeen kan geconcludeerd worden, 1. dat er doorgaans geen beduidende verschillen bestaan tussen informatie verkregen via een computer of via een schriftelijke versie; 2. dat computer-afnamen positiever gewaardeerd worden, zowel in vergelijking met een menselijke diagnosticus als met een schriftelijke vragenlijst; 3. dat geslacht en opleiding wellicht van invloed zijn op het omgaan met een computer; en 4. dat proefpersonen meer angst rapporteren voorafgaand aan een computer-afname in vergelijking met een schriftelijke afname.

#### VRAAGSTELLING

Dit artikel beoogt in de eerste plaats een bijdrage te leveren aan het equivalentie-onderzoek dat een noodzakelijke voorwaarde vormt voor het gebruik van gecomputeriseerde tests in de diagnostische praktijk. De tweede vraag heeft betrekking op de wijze waarop proefpersonen de computer als assessment-instrument beoordelen.

#### METHODE

##### *Proefpersonen*

Aan dit onderzoek verleenden 53 psychologiestudenten vrijwillig hun medewerking, te weten 30 vrouwen en 23 mannen. Hun gemiddelde leeftijd was 26 jaar ( $SD = 6.7$ ) en zij waren gemiddeld 4 jaar ingeschreven als student. Zeventien procent meldde geen enkele ervaring met computers te hebben, de overigen hadden in het kader van de studie één of meer cursussen gevolgd, drie van hen beschikten thuis over een PC.

##### *Materiaal*

*Computerprogramma.* De computerversie van de hieronder vermelde vragenlijsten werden afgenomen met behulp van het raamprogramma GAME (Groningen Advanced Measurement Environment; Brokken, 1986). Dit is een programma waarin de psycholoog-gebruiker zelfstandig zowel items als respons-categorieën kan definiëren in een eigen lay-out. Elk item wordt afzonderlijk aangeboden en

gaat vergezeld van één responsecategorie. Met behulp van een willekeurige toets (wij raadden het gebruik van de spatiebalk aan) wordt de cursor bij het gewenste antwoordalternatief geplaatst. (Dit betekent dus dat er geen 'verkeerde' toetsen zijn waarmee de onderzochte het programma in de war kan sturen.) Een druk op de 'RETURN'-toets voert het antwoord in en roept het volgende item op. In de huidige versie van GAME was het niet mogelijk om een eenmaal gegeven antwoord te corrigeren.

Voor de eigenlijke afname gebruikten we twee Olivetti's M24 met een monochroom (groen) 12 inch beeldscherm en een standaard IBM PC-toetsenbord. De items werden weergegeven tegen een donkere achtergrond. Een korte introductie op het scherm en enkele voorbeeld-items waren geïnstalleerd om de proefpersonen vertrouwd te maken met het toetsenbord en de responsemogelijkheden.

*Instrumenten.* Allereerst vulden de proefpersonen een korte *vragenlijst* in, waarop geïnformeerd werd naar leeftijd, geslacht, opleiding, jaar waarin men is begonnen met studeren, ervaring met computers tot nu toe, en de affiniteit tot het werken met computers.

Vier *Semantische Differentialen* (SD) met betrekking tot de woorden 'Computer', 'Radio', 'Testformulier' en 'Boek' werden schriftelijk afgenomen. De eerste twee begrippen kunnen beschouwd worden als de technologische pendanten van de laatste twee. 'Radio' en 'Boek' werden opgenomen als (verondersteld) neutrale woorden. De SD bedoelt een beeld te geven van de connotatieve betekenis van de target- en de neutrale woorden. De *a priori* dimensies 'Evaluatie', 'Activiteit' en 'Potentie' werden elk gerepresenteerd door vijf paren tegengestelde adjectieven, ontleend aan Sherry & Piotrowski (1986). De scoring van ieder adjectievenpaar vond plaats op een zevenpunts schaal. Per begrip dienden de proefpersonen dus  $5 \times 3 = 15$  adjectievenparen te scoren. Om de invloed van de testafname op deze connotaties te onderzoeken werd de SD twee keer afgenomen, de eerste keer *voor* de testafnamen en de tweede keer *erna*.

*Hopkins Symptom Check List* (HSCL; Luteijn et al., 1984) is een klachtenlijst bestaande uit 57 items die verwijzen naar psychisch en somatisch onwelbevinden. De items worden op een vierpunts schaal gescoord van 'Helemaal niet' (0) tot 'Heel erg' (3). Er kunnen drie scores berekend worden, namelijk de PSYCH-score (psychisch onwelbevinden, 17 items), de SOM-score (somatisch onwelbevinden, 8 items) en de TOTAAL-score die psychoneuroticisme meet. De referentieperiode is 'de voorafgaande week met vandaag erbij'. De HSCL werd zowel op de standaard schriftelijke (PP) wijze, alsook per computer (PC) afgenomen. De oorspronkelijke instructie werd op het scherm gepresenteerd.

De *T-schaal van de Amsterdamse Biografische Vragenlijst* (ABV; Wilde, 1970) meet testdefensiviteit, dat wil zeggen de mate waarin een proefpersoon bereid is openhartig op testvragen te antwoorden. De 23 items worden gescoord in

de categorieën 'juist', '?' en 'onjuist'. Voor het berekenen van de schaalscore worden de items gewogen opgeteld. De T-schaal is afgenomen om na te gaan of de twee methoden (PP en PC) verschil opleveren in de mate van defensiviteit als persoonlijkheidskenmerk. De instructie nodigt uit tot het scoren hoe men zich *in het algemeen* gedraagt. Ook de ABV-T-schaal werd zowel in de PP- als in de PC-uitvoering afgenomen. Ook in dit geval werd de oorspronkelijke instructie op het scherm gepresenteerd.

### Opzet

Per proefpersoon is een computer- (PC) en een 'paper and pencil'- (PP) afname van de HSCL en de ABV-T-schaal gerealiseerd. Om te controleren voor een mogelijk volgorde-effect werd een counter-balanced design toegepast met random toewijzing aan één van beide condities. Figuur 1 geeft dit schematisch weer. In beide condities vulden de proefpersonen de Semantische Differentialen vóór de eerste en na de tweede afname in.

De testafnamen vonden op dezelfde dag plaats met minstens vijf uur tussentijd. De 'paper-and-pencil'-afname geschiedde in kleine groepjes of individueel, terwijl de PC-afname altijd individueel was. Als globale uitleg voor het onderzoek werd de proefpersonen mondeling meegedeeld dat we onderzochten of men tests per computer zou kunnen afnemen. De proefleider bevond zich op discrete afstand om eventuele vragen te beantwoorden; een enkele maal werd van deze mogelijkheid gebruik gemaakt. De proefgroepen in beide condities verschilden niet van elkaar wat betreft hun samenstelling naar leeftijd ( $t = -0.43$ ,  $df = 1$ ; ns), geslacht ( $\chi^2 = 2.4$ ,  $df = 1$ ; ns) en voorafgaande ervaring met computers ( $\chi^2 = 4.1$ ,  $df = 4$ ; ns).

		METINGEN			
		PRE		POST	
Conditie I (n = 27)	SD	PP HSCL & ABV	PC HSCL & ABV	SD	12 vrouwen 15 mannen M = 25.7 jaar
	SD	PC HSCL & ABV	PP HSCL & ABV	SD	
Conditie II (n = 26)	SD	PC HSCL & ABV	PP HSCL & ABV	SD	18 vrouwen 8 mannen M = 26.5 jaar
	SD	PP HSCL & ABV	PC HSCL & ABV	SD	

Figuur 1. Onderzoeksdesign in schema.

*Noot.* PP: 'Paper-and-Pencil'; PC: Personal Computer; SD: Semantische Differentiaal; HSCL: Hopkins Symptom Checklist; ABV: Amsterdamse Biografische Vragenlijst; M: Gemiddelde.

## RESULTATEN

Gezien het relatief kleine aantal proefpersonen zijn de gebruikte statistische procedures non-parametrisch van aard. Er is op 1% significantieniveau getoetst, tenzij anders aangegeven.

*Equivalentie*

In tabel 1 zijn de gemiddelde scores op de T-schaal van de ABV en op de drie HSCL-schalen in beide condities weergegeven. Mann-Whitney-toetsen laten zien dat de proefpersonen in de tweede conditie systematisch (maar niet significant;  $p > .01$ ) hoger scoren op de HSCL-schalen, terwijl de gemiddelde ABV-T-score in de eerste conditie iets (maar niet significant;  $p > .01$ ) hoger ligt. De scores van de PP-

Tabel 1. Gemiddelden en standaarddeviaties van de scores op de ABV- en HSCL-schalen.

Conditie	Afname 1		Afname 2	
	PP		PC	
I (n = 27)				
ABV-T	29.8	(8.1)	30.0	(8.6)
HSCL-TOTAAL	14.9	(13.3)	14.1	(11.8)
HSCL-PSYCH	4.5	(5.3)	4.3	(4.6)
HSCL-SOMAT	1.3	(1.4)	1.4	(1.4)
II (n = 26)				
ABV-T	27.8	(6.7)	27.1	(7.0)
HSCL-TOTAAL	21.0	(13.8)	20.9	(14.6)
HSCL-PSYCH	7.1	(5.3)	7.1	(5.4)
HSCL-SOMAT	2.1	(2.3)	1.9	(2.2)

Noot: PP 'Paper-and-Pencil' afname; PC: Afname personal computer. Standaarddeviaties staan tussen haken.

Tabel 2. Verdelingseigenschappen van de HSCL en de ABV-T-schaal.

		M	SD	Kurtosis	Scheefheid
HSCL-TOTAAL	PP	17.8	14.2	1.19	1.25
	PC	17.5	13.1	1.00	1.07
HSCL-PSYCH	PP	5.7	5.5	0.28	1.02
	PC	5.7	5.1	0.12	0.85
HSCL-SOMAT	PP	1.6	1.9	3.69	1.79
	PC	1.7	2.0	4.07	1.88
ABV-T	PP	28.5	7.6	0.75	0.93
	PC	28.9	7.7	0.58	0.91

Noot: PP 'Paper-and-Pencil' afname; PC: Afname per personal computer. M: gemiddelde; SD: standaarddeviatie.

afnamen van beide condities kunnen derhalve tezamen genomen worden voor de verdere berekeningen; hetzelfde geldt voor de beide PC-afnamen.

Wanneer de scores per methode samengevoegd worden, levert dat de gemiddelden op die weergegeven zijn in tabel 2. Uit de tabel blijkt dat de PP- en PC-afnamen (zeer) sterk op elkaar gelijken ten aanzien van hun verdelingsparameters.

Om na te gaan in hoeverre samenhang tussen schalen zich wijzigt onder invloed van verschillende methoden, zijn Spearman rangcorrelaties berekend, zowel binnen alsook tussen de twee afnamemethoden (zie tabel 3).

De driehoeken linksboven en rechtsonder in de tabel betreffen correlaties binnen één methode; de rechthoek linksonder bevat de correlaties per schaal tussen de beide methoden. Wat het correlatiepatroon en de alfa-coëfficiënten betreft vertonen beide methoden grote overeenkomsten. Laatstgenoemde waarden zijn van dezelfde orde als die in de handleiding van de HSCL genoemd worden (Luteijn et al., 1984). Op de diagonaal in de onderrechtshoek staan de (cursief gedrukte) correlaties tussen beide condities per schaal. Voor de ABV-T-schaal en voor de SOMAT-schaal van de HSCL geldt dat de correlaties bevredigend hoog zijn. De totaalscore en de PSYCH-schaal van de HSCL vertonen correlaties die overeenkomen met de stabiliteitscoëfficiënten zoals vermeld in de handleiding (Luteijn et al., 1984).

*Evaluatie*

Voorafgaand aan de testafnamen werd op een vijfpuntschaal (lopend van 'Erg vervelend' tot 'Erg leuk') gevraagd hoe men het werken met een PC waardeerde. De mediane score was 4 (= 'Leuk'), waarbij 75,6% de waardering '(erg) leuk' aankruiste.

Bij de drie *a priori* dimensies van de Semantische Differentiaal zijn de gemiddelde itemscores berekend (range 1 tot 7), waarbij een hogere score duidt op een sterkere aanwe-

Tabel 3. Spearman-correlaties tussen ABV-T en HSCL-schalen en hun coëfficiënten alfa.

		PP				PC			
		ABV-T	HSCL	PSYCH	SOMAT	ABV-T	HSCL	PSYCH	SOMAT
PP	ABV-T	.75							
	HSCL	-.15	.91						
	-PSYCH	-.26	.93	.86					
	-SOMAT	.03	.38	.24	.67				
PC	ABV-T	.87	-.19	-.32	.04	.74			
	HSCL	-.33	.75	.72	.41	-.29	.91		
	-PSYCH	-.40	.68	.74	.32	-.36	.94	.86	
	-SOMAT	.04	.37	.23	.80	.04	.52	.34	.69

Noot: PP 'Paper-and-Pencil' afname; PC: Afname per computer. Decimale punten zijn weggelaten. Op de hoofd diagonaal staan (cursief gedrukt) de alfa-coëfficiënten vermeld. Bij correlaties groter dan .33 is  $p < .01$ .

Tabel 4. Item-gemiddelden en standaarddeviaties van de Semantische Differentiaal bij de pre- en posttest.

Begrip	Dimensie	Pretest		Posttest	
		M	SD	M	SD
'Computer'	Evaluatie	3.4	0.7	3.4	0.7
	Potentie	3.4	0.6	3.4	0.8
	Activiteit	3.1	0.7	3.3	0.7
'Testformulier'	Evaluatie	3.0	0.6	2.9	0.6
	Potentie	2.6	0.8	2.7	0.7
	Activiteit	2.6	0.7	2.6	0.7
'Boek'	Evaluatie	4.3	0.7	4.2	0.7
	Potentie	3.2	0.7	3.3	0.7
	Activiteit	3.5	0.8	3.5	0.7
'Radio'	Evaluatie	3.8	0.8	3.8	0.8
	Potentie	3.0	0.6	3.1	0.6
	Activiteit	4.0	0.8	4.0	0.7

Noot: M: gemiddelde; SD: standaarddeviatie.

zigheid van de positieve pool van de desbetreffende dimensie. In tabel 4 staan de gemiddelden en standaarddeviaties van de pre- en posttest. Aan de scores is te zien dat de connotaties niet veranderen onder invloed van deelname aan het onderzoek.

Om na te gaan of de targetwoorden 'Computer' en 'Testformulier' van elkaar verschillen, is per semantische dimensie een Wilcoxon-test uitgevoerd. Het bleek dat op elk van de drie dimensies en bij beide afnamen het begrip 'Computer' een significant ( $P < .001$ ) hogere score had dan het begrip 'Testformulier'. Dit duidt erop dat de proefpersonen een positievere waardering, grotere potentie en meer activiteit aan het begrip 'Computer' toekenden.

## DISCUSSIE

De resultaten van het onderzoek laten zien dat een schriftelijke en een computer-afname van de Hopkins Symptom Checklist equivalent zijn. Dit blijkt uit:

- de afwezigheid van significante verschillen in gemiddelden tussen beide afnamemethoden;
- bevredigend hoge correlaties *tussen* de twee afnamemethoden;
- overeenkomstige correlatiepatronen *binnen* de twee methoden.

Aan de eisen van de American Psychological Association inzake equivalentie van test scores is met de eerste twee punten voldaan.

Een opvallend punt is het verschil in HSCL-scores van de proefpersonen in de beide condities (zie tabel 1). Hiervoor is niet onmiddellijk een verklaring te geven. Vermoedelijk hebben we te maken met een toevalseffect, waardoor er iets meer hoogscorers in de tweede conditie terecht zijn gekomen (vrouwen scoren doorgaans wat hoger op klachtenlijsten dan mannen; zij zijn in conditie II wat sterker vertegenwoordigd). Onze verwachting is dat dergelijke verschillen bij voldoende grote proefgroepen opgeheven worden.

Een tweede punt is, dat volgens de Semantische-Differentiaal-uitkomsten de computer positief gewaardeerd werd in vergelijking met het testformulier. Dit zou veroorzaakt kunnen worden doordat onze proefpersonen een zekere affiniteit en ervaring met computers hadden en het werken ermee als interessant beschouwden. Onze proefpersonen - jonge, goed opgeleide mensen - zijn niet representatief voor de bevolking in het algemeen. De vraag blijft derhalve bestaan, hoe een minder ingewijde groep een computer-afname ervaart. Het zou daarom onverstandig zijn om onze resultaten zonder meer naar andere groepen te generaliseren. Mathisen et al. (1985) vonden bijvoorbeeld dat vrouwen, ouderen, mensen met een lage opleiding en met weinig computerervaring moeite hadden om de testafname

per computer zelfstandig en met succes tot een goed einde te brengen. Van de Vijver (1987, p. 14) stelt zelfs; 'Computer-ondersteunde tests meten namelijk ook het vermogen om met dergelijke apparatuur om te gaan'. Dit blijkt bij *speed*- en intelligentietests zeer duidelijk met de testperformance te interfereren.

Enkele opmerkingen over equivalentie-onderzoek in het algemeen zijn hier op hun plaats. De relatief korte tijd tussen beide afnamen kan een zeker herinneringseffect te weegbrengen. Eleganter zou het zijn om in dergelijke gevallen te werken met paralleltests of met een langere periode tussen de afnamen. Verder zou men aan de twee door ons gebruikte condities nog een tweetal kunnen toevoegen, namelijk een afname op beide tijdstippen van alleen de schriftelijke versie, alsook een afname van alleen de computer-versie; dit zijn dus hertests in de eigenlijke zin van het woord.

Equivalentiestudies zijn sterk 'device'-gebonden, d.w.z. dat ze nauw samenhangen met de specifieke computerconfiguratie en de gebruikte software. Onderzoekers dienen tot het (op het oog zeer banale en alledaagse) niveau van de concrete afnamesituatie af te dalen. Uitwisseling van informatie over procedures die men gevolgd heeft, kan inzichtelijk maken waarom er overeenkomsten en verschillen tussen studies bestaan. Juist omdat er in Nederland (en ook daarbuiten) wat dat betreft geen enkele standaard voorhanden is voor het presenteren van tests op een pc of mini-computer, dienen onderzoeksverslagen in detail te treden over de afnamesituatie en de gebruikte computerconfiguratie. Men kan denken aan het vermelden van het soort toetsenbord of 'muis' dat gebruikt is en aan andere kenmerken van de item-presentatie. Het door ons gebruikte programma *GAME* is bijvoorbeeld in staat om zeer snelle beeldwisselingen uit te voeren. In de praktijk blijkt dat sommige proefpersonen deze snelheid als storend ervaren, zodat nu overwogen wordt om een vertraging tussen schermovergangen in te bouwen. Soortgelijke problemen kunnen optreden wanneer dezelfde programmatuur op pc's met verschillende kloksnelheden geïnstalleerd is (10 MegaHerz in plaats van 4.77 MegaHerz). Een belangrijk punt is tenslotte de vraag, of de hard- en software idiot-proof is (als superlatief van fool-proof); zelfs de onhandigste proefpersoon zou geen brokken moeten kunnen maken.

De geringe beschikbaarheid van diagnostische programmatuur (en zeker van schil- of raamprogramma's) kan in zekere zin positief geïdentificeerd worden. Voordat de markt overspoeld wordt met allerhande producten, kunnen er voorstellen voor standaardisering en voor kwaliteitseisen geformuleerd worden. Enkele daarvan zijn reeds op schrift gesteld door de American Psychological Association (1986). Zij geeft een aantal richtlijnen voor het afnemen van tests per computer, zowel voor de eindgebruiker als voor de testontwikkelaar. In essentie komt het erop neer dat de afnamesituatie zo optimaal mogelijk (dat wil zeggen rustig en zonder technische storingen) moet zijn voor de geteste per-

soon, zodat deze zich ongestoord aan de opdrachten kan wijden. Het verschaffen van oefen- en kennismakingsfaciliteiten binnen het programma maakt daarvan een belangrijk deel uit. Onze proefpersonen bleken hierdoor zeer snel te begrijpen wat er van hen verlangd werd en hoe zij een respons konden invoeren. De APA (1986) beveelt verder aan om 'terugbladeren' naar eerder gegeven responsen mogelijk te maken, met de additionele mogelijkheid van correctie; in al deze opzichten dient een geautomatiseerde test niet af te wijken van een schriftelijke test. We zouden daar nog aan toe willen voegen, dat het uit oogpunt van transporteerbaarheid van gegevens aanbeveling verdient, dat een programma de mogelijkheid biedt om de verzamelde scores weg te schrijven in zogenaamde ASCII-code die door vrijwel alle computers 'leesbaar' is.

Tenslotte verdient het aanbeveling dat een toezichhoudend orgaan van deskundigen (zoals bijvoorbeeld de COTAN; cf. Akkerman & Takens, 1984; Koning, 1985) kwaliteitsbeoordelingen van computerproducten publiceert. Dit zou zowel moeten gelden voor raamprogramma's als voor 'kant-en-klare' gecomputeriseerde tests. Enkele richtlijnen daarvoor zijn reeds geformuleerd (Koning, 1985). Met duidelijke aanbevelingen en richtlijnen zou een belangrijke stap gezet zijn in de richting van meer eenduidigheid en vergelijkbaarheid van testprogramma's en geautomatiseerde procedures. Voor de modale gebruiker wordt het dan mogelijk om het kaf van het koren te scheiden. We delen wat dat betreft niet op voorhand de kritiek van Akkerman (1986) op 'lege schil'-programma's zoals CDL (Computer Diagnostiek Leiden; Rombouts, 1986). Zijn argument is dat gebruikers eigen (ongevalideerde) versies kunnen maken, en dat het sommigen aan deskundigheid zou kunnen ontbreken om met een dergelijk systeem om te gaan. Dit gevaar is echter niet specifiek voor gecomputeriseerd testgebruik, maar geldt evenzeer voor welke vorm van psychodiagnostiek dan ook. In principe kan ieder stuk onbeschreven papier voor een soortgelijk vergrijp misbruikt worden. Wat het diagnostische medium ook is, papier, computer, gesprek of observatie, steeds is het aan de psycholoog om daar verantwoord mee om te gaan.

## LITERATUUR

- Akkerman, A.E. (1986). Liever een lege schil? Reactie op CDL. *Psychologie & Computers*, 3, 71-72.
- Akkerman, A.E. & Takens, R.J. (1984). Computer-ondersteunde psychodiagnostiek in Nederland. Rapport van de Taakgroep COP. *De Psycholoog*, XIX, 145-156.
- American Psychological Association (1986). *Guidelines for computer-based tests and interpretation*. Washington. APA.
- Angle, H.V., Ellingwood, E.H., Hay, W.M., Johnsen, T., & Hay, L.R. (1977). Computer-aided interviewing in comprehensive behavioral assessment. *Behavior Therapy*, 8, 747-754.
- Brokken, F. (1986). *GAME Groningen Measurement Environment*. Raamprogramma onder MS-DOS.
- Butcher, J.N., Keller, L.S., & Bacon, S.F. (1985). Current de-

- velopment and future directions in computerized personality assessment. *Journal of Consulting and Clinical Psychology*, 53, 803-815.
- Cruickshank, P J (1982). Patient stress and the computer in the consulting room. *Social Science and Medicine*, 16, 1371-1376
  - Greist, J.H., Gustafson, D.H., Stauss, F., Rowse, G L., Laughren, T P., & Chiles, J A. (1973) A computer interview for suicide risk patients. *American Journal of Psychiatry*, 130, 1327-1332
  - Harrell, T.H., & Lombardo, T A. (1984) Validated of an automated 16PF administration procedure. *Journal of Personality Assessment*, 48, 638-642.
  - Hermans, H.J.M., Hermans-Jansen, E., & Gilst, W van (1985). Does the medium change the message? *De Psycholoog*, XX, 122-125
  - Johnson, J.H., & Williams, T A. (1975). The use of on-line computer technology in a mental health admitting system. *American Psychology*, 30, 388-390.
  - Katz, L., & Dalby, J.T (1981). Computer and manual administration of the Eysenck Personality Inventory *Journal of Clinical Psychology*, 32, 586-588.
  - Knol, D L., & Schoonman, W (1986). De constructie van een gecomputeriseerde capaciteitstest met gebruikmaking van responstijden. In: W.J. van der Linden (Red.), *Moderne methoden van testconstructie en -gebruik*. Lisse: Swets & Zeitlinger.
  - Koning, R.F. (1985) De beoordeling van computer-ondersteunde tests. *De Psycholoog*, XX, 561-565.
  - Koning, R.F (1986). Ontwikkelingen in de computer-ondersteunde psychodiagnostiek. In: G. Kempen & R J. Takens (Red ), *Psychologie. Informatica en Informatisering* Amsterdam. NIP.
  - Lucas, R.W. (1977). A study of patient's attitudes to computer interrogation. *International Journal of Man-Machine Studies*, 9, 69-86
  - Lushene, R.E., O'Neil, H.F , & Dunn, T (1974). Equivalent validity of a completely computerized MMPI. *Journal of Personality Assessment*, 38, 353-361.
  - Luteijn, F., Hamel, L., Bouman, T.K., & Kok, A R. (1984) *HSC L. Hopkins Symptom Check List*. Lisse: Swets & Zeitlinger.
  - Meier, S T. (1985) Computer aversion. *Computers in human behavior*, 1, 171-179.
  - Mathisen, K S , Evans, F J , Meyers, K., & Kogan, L. (1985). Human factors influencing patient-computer interaction *Computers in human behavior*, 1, 163-170.
  - Reinders, A , & Van Lenthe, J (1986). *TL - Een programmeertaal voor testkonstruktors*. Heymans Bulletin Psychologische Instituten RUG. HB-86-814-RP.
  - Rombouts, R., & Diekstra, R F. (1987). Computer Diagnostiek Leiden (CDL) *De Psycholoog*, XXII, 16-18
  - Sherry, D., & Piotrowski, C. (1986) Consistency of factor structure of the semantic differential: An analysis of three adult samples. *Educational and Psychological Measurement*, 46
  - Skinner, H.A , & Allen, B.A. (1983). Does the computer make a difference? Computerized versus face-to-face versus self-report assessment of alcohol, drug, and tobacco use. *Journal of Consulting and Clinical Psychology*, 51, 267-275.
  - Vliet, W.J D van & Bonarius, J.C.J. (1984). Psychodiagnostiek en computers *De Psycholoog*, XIX, 110-118
  - Vijver, F. van de (1987). Computer-ondersteunde tests in de diagnostische praktijk. *De Psycholoog*, XXII, 10-15.
  - Wagenborg, J.E.A., & Koning, R.F. (1984). De Hopkins Symptom Check List geautomatiseerd. *Psychologie en Computers*, 1, 18-21
  - White, D M., Clements, C.B., & Fowler, R.D. (1985). A comparison of computer administration with standard administration of the MMPI *Computers in human behavior*, 1, 153-162.
  - Wilde, G.J. (1970). *Neurotische labiliteit gemeten volgens de vragenlijstmethode*. Amsterdam: Van Rossen.
  - Zaal, J. (1987). Testen achter de beeldbuis *De Psycholoog*, XXII, 19-24.