



Universiteit
Leiden
The Netherlands

Tien? Reactie op F. J. Heunks, Tien kleine vergissingen en één grote; AP XI, p. 406

Boon, M.E.; Niemöller, B.

Citation

Boon, M. E., & Niemöller, B. (1976). Tien?: Reactie op F. J. Heunks, Tien kleine vergissingen en één grote; AP XI, p. 406. *Acta Politica*, 11: 1976(3), 408-419. Retrieved from <https://hdl.handle.net/1887/3451637>

Version: Publisher's Version

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/3451637>

Note: To cite this publication please use the final published version (if applicable).

De konklusie moet dan zijn dat men ondanks aanwijsbare ongunstige omstandigheden bij de dataset van het hoofdonderzoek, de schaal toch betrekkelijk veilig kan gebruiken bij de analyse van die dataset. De samenhangen die in het onderzoeksmodel worden verwacht, zullen echter waarschijnlijk wat lager uitvallen. Tegelijk dient men de lezer er op te wijzen dat hij bij verdere toepassingen ook de nodige aandacht geeft aan de items die in het vooronderzoek in de schaal worden opgenomen. Dit alles kan men lezen in mijn proefschrift. Schaalontwikkeling is een langdurige en veelzijdige zaak, en bepaald ingewikkelder dan het automatisme van een bepaalde rekenprocedure.

Literatuur

- Green, B. F., A method of scalogram analysis using summary statistics', in: *Psychometrika*, Vol. 21, 1956, pp. 79-88
 Heunks, F. J., *Aliëntatie en stemgedrag*, Tilburg, 1973.
 Mokken, R. J., *A theory and procedure of scale analysis*, Den Haag, 1970

Tien?

Reaktie op F. J. Heunks 'Tien kleine vergissingen en één grote'

door M. E. Boon en B. Niemöller

Laten we Heunks' verslag van de ontwikkeling van zijn aliëntatieschalen eens op de voet volgen om na te gaan of de 'leerlingen van Mokken' wel zulke simpele zielen zijn als Heunks suggereert. Wellicht wordt dan tevens duidelijker dan in ons eerste beknopte stuk waarom we vinden dat de tien schalen van Heunks nog lang niet allen voor algemeen gebruik geschikt zijn.

1

Heunks noemt in het begin van zijn verantwoording twee mogelijke werkwijzen om tot schaalconstructie te komen, namelijk de exploratieve – en de toetsende – werkwijze. Zijn besluit om voor de toetsende werkwijze te opteren rechtvaardigt hij als volgt: 'Wij meenden theoretisch sterk genoeg te staan om deze laatste weg aanstonds te bewandelen. Begrippen als machteloosheid zijn theoretisch duidelijk plaatsbaar, men zie daarvoor de laatste paragrafen van hoofdstuk 1 en 2. Voor de volledigheid is in een latere fase de alternatieve werkwijze eveneens nog gevolgd.' (Heunks, p. 263)

Of de genoemde *begrippen* inderdaad zo duidelijk plaatsbaar zijn lijkt ons op dat punt van zijn betoog minder relevant. Het gaat erom of de door Heunks gekozen items goede *indikatoren* zijn voor die begrippen. Dat later 'voor de volledigheid' de alternatieve werkwijze nog wordt gevolgd lijkt ons een minder juiste volgorde.

2

Het toetsen van zijn a priori schalen doet Heunks m.b.v. Green's Index of Consistency' waarover hij o.a. opmerkt: 'Deze "Index of Consistency" is behalve op

de gehele groep van schaalitems ook toepasbaar op itemparen daaruit, hetgeen in staat stelt tot het selecteren van die items welke tesamen de beste schaalbaarheid vertonen.' (Heunks, p. 264)

Dit lijkt ons om een aantal redenen een merkwaardige uitspraak:

- Uit de indices voor item-paren valt niets met zekerheid te zeggen omtrent de index voor een eventuele combinatie van meer dan twee items. Dit vormt een van de problemen die men bij exploratieve schaalanalyse tegenkomt.
- Heunks wijst later (Heunks, p. 266 voetnoot 2 en Heunks' reaktie punt 6) dit soort zoek procedures af.
- Bestudering van de matrices met Green's I voor alle itemparen (Heunks, pp. 388/9) zou het zelfvertrouwen van Heunks ten aanzien van zijn theoretische a priori-schalen moeten ondermijnen. (zie verder punt 6)

De vraag rijst nu hoe serieus Heunks de mogelijkheden van de door hem gekozen index heeft benut.

3

Het onder punt 2 te berde gebrachte heeft overigens niets te maken met onze waardering voor Green's index I. Integendeel, voor Green's I geldt, evenals voor Jackson's PPR en Loevinger's H, dat zij '... are all objective, and result in measures which are not affected by the distribution of item difficulties.' (White and Saltz, p. 189) Aangezien wij in het navolgende gebruik zullen maken van Loevinger's H lijkt het volgende citaat nog op zijn plaats.

'All have the same underlying rationale, but differ slightly in the way in which "errors" are counted. Loevinger's Ht is the most conservative of the three since all possible errors are counted; Jackson's PPRt is the least conservative, and Green's I will usually fall between the two.' (White and Saltz, p. 189)¹

4

Alvorens tot de presentatie van de met behulp van Green's I gemeten kwaliteit van zijn a priori schalen over te gaan, geeft Heunks nog een toelichting m.b.t. de aan I te stellen 'ondergrens'.

'Een I van omstreeks .50 wordt door Green bevredigend geacht, terwijl een lagere I tot .25 toe twijfelachtig wordt geacht.

Nog lagere I-waarden duiden volgens hem op het ontbreken van schaalbaarheid in een Guttman-zin.' (Heunks, p. 264)

Deze aanduidingen moeten worden opgevat als empiriese bevindingen; een theoretische basis ontbreekt. Hetzelfde geldt voor de vuistregel die Mokken geeft voor Loevinger's H:

- a – $0.50 \leq H$: a strong scale;
- b – $0.40 \leq H < 0.50$: a medium scale;
- c – $0.30 \leq H < 0.40$: a weak scale.

(Mokken, 1971, p. 185)

Wel kunnen we hieruit konkluderen dat men met item-kombinaties die een index hebben in de buurt van de .30 of zelfs .25 wel bijzonder voorzichtig dient te zijn als het om toekenning van het predikaat 'schaal' gaat.

Gelet op het onder punt 3 gestelde moeten we voorts konkluderen dat een ondergrens voor Green's I als regel *hoger* gesteld moet worden dan die voor Loevinger's H.

5

En dan nu de resultaten van Heunks toetsing van zijn tien a priori schalen.

Tabel 1:

Overzicht van de schaalbaarheid der items in het hoofdonderzoek ontleend aan Heunks p. 265 en bijlage I)

schaal ²	Green's I	schaal ²	Green's I
EMA (52, 61, 62, 63)	.47	MMA (2, 6, 8, 10)	.35
EMS (65, 69, 70, 71)	.35	MMS (16, 19, 20, 21)	.24
ENE (74, 77, 78, 80)	.36	MN (22, 24, 26, 27)	.32
ENK (81, 82, 83) ³	.16	MD (30, 32, 33, 34)	.25
ED (86, 87, 88, 90)	.32		
EA (92, 93, 94, 97)	.58		

Gelet op de in punt 4 aangehaalde waarderings richtlijn (I tot .25) leidt tabel 1 in ieder geval tot de konklusie dat we bij de a priori 'schalen' MMS, MD en ENK toch slecht van schaalbaarheid kunnen spreken.

Voor de overige schalen, met uitzondering van EA, dient de schaalbaarheid, alweer volgens het citaat uit Heunks, twijfelachtig te worden genoemd.

De door Heunks gerapporteerde I's komen aardig overeen met de resultaten welke wij verkregen met behulp van een toetsende schaalanalyse volgens Mokken⁴ (Mokken, 1971, H.5). Het daarbij gebruikte schaalbaarheidskriterium, de eerder genoemde H van Loevinger, volgde Green's I vrij nauwkeurig, zij het meestal wat lager (let op ons punt 4). Uitzondering vormde de schaal ENE: wij vonden een H van .24 waar Heunks een I van .36 rapporteert. Gelet op de Green indices voor de item-paren uit deze schaal (Heunks, bijlage 6, p. 388) lijkt die .36 ons niet erg betrouwbaar.

In verband met punt 9 uit Heunk's kritiek voor alle duidelijkheid: we gaan er vanuit dat de ENE-schaal uit de items 74, 77, 78 en 80 bestaat. De H-koëfficiënt voor alle item-paren geven we in tabel 2 weer.

Het lijkt waarschijnlijk dat de .63 voor paar (74-80) in Green's Index de kwaliteit van de hele schaal te veel flatteert.

6

Blijven we nog even bij de inter-item I van Green zoals die voor de items uit het hoofdonderzoek in bijlage 6 en 7 van Heunk's proefschrift zijn opgenomen.

Bestudering daarvan zullen we aan de hand van onze weergave van genoemde bijlagen doen.

We hebben daartoe in tabel 3 en 4 de inter-item I van Green voor alle paren van

Tabel 2:

Loevinger's	Hij voor ENE item-paren			
	74	77	78	80
74	—			
77	.17	—		
78	.18	.31	—	
80	.63	.23	.16	—

items welke in een schaal zijn opgenomen, weergegeven voorzover ze boven de door Green gestelde .2-5 uitkomen.

We geven daarbij tevens aan in welke schaal Heunks de items a priori geplaatst heeft.

Het valt onmiddellijk op dat de samenhang tussen een item uit een bepaalde a priori schaal en alle items uit een andere a priori schaal vaak opmerkelijk hoog is. Met name valt dat op voor ED-items t.o.v. EMA en EMS items; voor EMA-items t.o.v. ED en EMS items; en EMS-items t.o.v. ED en EMA items. Ook voor de maatschappelijke items valt iets dergelijks waar te nemen zoals uit tabel 4 blijkt.

We herinneren hier aan een opmerking uit punt 2: we kunnen niet zonder meer tot bepaalde schalen konkluderen op grond van deze indices voor paren. Wel kunnen we ons gaan afvragen in hoeverre bijvoorbeeld ED, EMA en EMS items tot te onderscheiden dimensies behoren.

7

Waar Heunks ondanks zijn grote kennis van de betrokken begrippen en mogelijke indicatoren daarvan ons inziens toch nog ten onrechte voor de toetsende schaalanalyse koos, kunnen wij niet anders dan voor een eksploratieve benadering opteren. Dit stelt ons dan tevens in staat te controleren voor '... alternatieve referentiekaders van de respondenten in vergelijking met de aan de literatuur ontleende kaders.' (Heunks kritiek, punt 6)

Uitgaande van de verzameling elektorale items zochten wij m.b.v. Mokken-schaalanalyse (zoekprocedure)⁵ naar daarin aanwezige schalen. Zoals we al opmerkten (punt 2) wijst Heunks erop '... dat door het selekteren van een kleine subset van items uit een groot aantal items, de schaalbaarheid bij toeval kan ontstaan.' (Heunks' kritiek, punt 6). In de door ons gevolgde werkwijze wordt gepoogd dit gevaar van 'capitalizing on chance' sterk terug te brengen door "a continual reduction of the level of confidence in consecutive steps." (Mokken, 1971, p. 196). Alhoewel het hier een min of meer arbitraire oplossing betreft spreekt Mokken de hoop uit dat '... these probably strong requirements concerning the confidence levels will mitigate the effects of manipulating chance results.' (Mokken, 1971, p. 197)

Het uiteindelijke resultaat voor de elektorale items is⁶:

'Schaal' 1 - 94 (EA), 93 (EA), 92 (EA), 97 (EA);

2 - 71 (EMS), 65 (EMS), 69 (EMS), 57 (EMA), 64 (EMA), 63 (EMA);
52 (EMA), 88 (ED) ...;

3 - 87 (ED), 52 (EMA), 61 (EMA), 62 (EMA), 59 (EMA), 70 (EMS);
86 (ED), 63 (EMA), 88 (ED) ...;

4 - 78 (ENE), 77 (ENE) (H=.31);

Tabel 3:

[illegible]

Tabel 4:

		MMA			MMS			MN			MD						
		2	6	8	10	16	19	20	21	22	24	26	27	30	32	33	34
MMA	2	—															
	6	.42	—														
	8	.33	.27	—													
	10	.35	.45	.34	—												
	16	.31	.33	.42	.30	—											
MMS	19	.35	.29	.30		.33	—										
	20		.29	.27		.26		—									
	21		.35			.26	.33		—								
	22		.42	.45	.30	.46				—							
	24	.33	.36	.39	.28	.52				.34	—						
MN	26			.26	.32	.28					.31	—					
	27				.44						.39	.30	—				
	30	.29	.42	.35	.33	.42				.31	.38		.29	—			
	32									.45	.44			.33	—		
	33									.47	.38			.32	.28	—	
MD	34	.42	.28	.38	.28	.28				.39	.38		.26	.49			—

(items 57, 64 en 59 zijn uiteindelijk niet in een schaal van het hoofdonderzoek opgenomen).

De 'schalen' 1, 2 en 3 hebben een Loevinger's H tussen de .40 en .60.

Voorts valt te vermelden dat 'schaal' 2 bij doorzetten van de zoekprocedure, uit te breiden valt met de items uit 'schaal' 3. Het is daarbij opvallend dat

– de volgorde waarin die items worden toegevoegd globaal gezien omgekeerd is aan de opname volgorde van die items in 'schaal' 3, dat wil zeggen schaal 2 kan worden uitgebreid met het item dat in schaal 3 als laatste wordt toegevoegd, etc.

– de schaalbaarheid loopt daarbij nauwelijks terug ($H = \pm .40$)

– hetzelfde geldt voor 'schaal' 3; uitbreiding is mogelijk met de items uit 'schaal' 2, ook weer in ongeveer omgekeerde volgorde.

Om deze dupliek niet al te uitgebreid te doen worden, laten we onze nadere analyse van de maatschappelijke items hier en in het navolgende verder achterwege.

8

We komen nu tot een voorlopige evaluatie van Heunk's toetsende bevindingen (hiervoor genoemd onder punt 5), onze ongerustheid t.a.v. de onderscheidbaarheid der diverse dimensies (punt 6) en de eigen exploratieve bevindingen (punt 7). In zijn eigen evaluatie konstateert Heunks dat de kwaliteit van zijn a priori schalen een stuk achterblijft bij de resultaten van zijn proefenquête. Hij geeft daarvoor als oorzaken:

a – Het hoofdonderzoek werd gekenmerkt door een relatief klein aantal beschikbare items per beoogde schaal, in verhouding tot het daarin minimaal op te nemen items.

b – Om in het hoofdonderzoek (dat i.t.t. de proefenquête moest dienen voor verschillende onderzoekingen) aan het minimum aantal van vier beschikbare items per beoogde schaal te komen, moest gebruik worden gemaakt van vervangende items uit een ander onderzoek. Deze, door Heunks verre van ideaal genoemde noodoplossing, leidde bijvoorbeeld voor de MMS-schaal tot een vervanging van drie van de vier items.

c – In het hoofdonderzoek lag de non-response hoger dan in de proefenquête (15% tegen 5%).

d – Alleen voor de ENK-schaal: te elfder ure werden item-teksten bijgeslepen (Heunks, pp. 266/8).

'Door dit alles' aldus Heunks, 'moet ten aanzien van de meeste schalen in het hoofdonderzoek worden gesproken van kwasi-schaalbaarheid, volgens het criterium van Green. De resultaten der proefonderzoekingen en de hierboven gegeven verklaring voor de geringe schaalbaarheid in de hoofdenquête, maken echter de één-dimensionaliteit van onze schalen weer minder betwistbaar. Met uitzondering van de ENK-schaal die zelfs het nivo van kwasi-schaalbaarheid niet haalde in het hoofdonderzoek, beschouwen we de schalen dan ook als éénduidige schalen en in zoverre geschikt als variabele bij de verdere analyse.' (Heunks, p. 268)

Wij vinden dat, in dit stadium, wel een bijzonder optimistische uitspraak. Immers, op grond van de resultaten van Heunks (tabel 1) kwamen wij in punt 5 tot de konklusie dat we bij de a priori 'schalen' MMS, MD en ENK echt niet meer van schaalbaarheid kunnen spreken, terwijl volgens eigen berekeningen ENE eveneens onvoldoende schaalkwaliteiten bezit.

Nu valt er bij dit soort arbitraire maten geen scherpe overgang van schaalbaar

naar niet-schaalbaar te markeren: een 'schaal' met een I van .25 is niet opeens veel slechter dan een schaal van .26. Gelet echter op het feit dat de I een optimistische maat is (zie punt 4) en de door Mokken ($H \geq .30$) en Green ($I \geq .26$) gegeven richtgetallen, menen wij dat Heunks' optimisme onvoldoende grond heeft. De vier door Heunks genoemde punten (zie boven punten a t/m d) kunnen daarbij niet dienen om 'de ééndimensionaliteit van onze schalen minder betwistbaar' te maken. Hoogstens kan men er voor een vervolgonderzoek enige moed uit peuren. Zoals aangekondigd (punt 1) heeft Heunks zijn toetsende werkwijze 'voor de volledigheid' nog laten volgen door een exploratieve werkwijze teneinde na te gaan of er mogelijk alternatieve schalen zijn. Daartoe paste hij op het materiaal van het hoofdonderzoek een non-metrische faktor-analyse toe, o.a. uitgaande van Yules Q voor item-scores. We citeren nu hoe dat voor de elektorale items in zijn werk ging.

'Begonnen is met een analyse van 30 elektorale items. Daarop werden echter met geen van de drie coëfficiënten goede resultaten behaald (de 'stress' bleef te hoog, ofwel de ruimtelijke benadering van de coëfficiënten bleef te grof). Na verwijdering van zeven items en toevoeging van een nieuw item werd evenmin een goed resultaat geboekt. Na verdere reductie van het aantal items tot 19 en vervanging van het vijf-factoren-model door het vier-factoren-model werd een wat beter resultaat bereikt, dat echter blijkens navolgend overzicht van 'ladingen' niet veel mogelijkheden geeft tot zinvolle interpretatie. Wel blijkt dat de resultaten eerder een ondersteuning bieden van de beoogde schalen dan aanleiding geven tot andere structuren tussen de items.' (Heunks, p. 273)

We zijn het volledig met Heunks eens dat er weinig mogelijkheden zijn tot zinvolle interpretatie. Ondersteuning van de beoogde schalen blijft beperkt tot schaal EA, terwijl onze vrees dat items uit bijvoorbeeld de schalen EMA en EMS nogal verwisselbaar zijn niet wordt weggelaten. (Voor de resultaten zie Heunks, p. 274.) Heunks is echter nog verder gegaan met zijn exploratie: op de inter-item I paste hij kluster-analyse toe. Hij koos daarvoor de kluster-analyse van Holzinger en Harman waarvan hij zegt:

'Deze techniek leidt tot klusters waarbinnen de samenhangen tussen de items relatief hoog zijn, terwijl de samenhangen van deze items met items buiten het kluster relatief laag zijn. Nadeel van deze techniek is het uitvoeren van berekeningen welke een interval-meetnivo veronderstellen, terwijl de onderzoeksgegevens en de gehanteerde maat van samenhang van ordinaal nivo zijn' (Heunks, p. 276, noot 1).

Zijn resultaten geven we weer in tabel 5.

Heunks signaleert een sterke overeenkomst met zijn schaalanalyse en is hiermede, niet ten onrechte, wel tevreden. Met klusters 1 en 4 kan hij inderdaad tevreden zijn, maar cluster 2 levert moeilijkheden op. We kunnen daar toch niet van een klustering van EMA items spreken want bij het eerste EMA item (52) worden een ENK (82) en een EMS (71) item in de kluster opgenomen. Het volgende opgenomen item, EMA-63 dankt zulks mede aan de zuigkracht van EMS-71: Green's I voor paar (63-71) is .64.

Uit cluster 3 blijkt nog eens hoe moeilijk EMS en EMA items uit elkaar te houden zijn. Dat de EMA-items 57 en 64 niet in een schaal werden opgenomen en EMS-item 71 door Heunks 'verdwaald' wordt genoemd, leidt niet tot de konklusie dat cluster 2 een EMA-kluster en cluster 3 een EMS-kluster kan worden genoemd. Daartoe zou met weglating van de met x-gemerkte items een nieuwe klusteranalyse

moeten worden gedaan, waarbij het resultaat zich slecht laat voorspellen.⁸

Wij hebben op dit punt ook nog wat exploratief werk gedaan, en wel een klusteranalyse volgens Johnson's hiërarchische klusteranalyse (interval-nivo niet vereist), maximum methode, op Yule's Q.

De resultaten daarvan kunnen we als volgt kort weergeven:

- 1 – er is een EA kluster;
- 2 – de EMA items 52, 61, 62 en 59, ED-items 86 en 87 en EMS-70 worden in hetzelfde kluster opgenomen;
- 3 – de EMS-items 71, 69 en 65, de EMA-items 63, 64 en 57 en ED-88 vormen samen een kluster;
- 4 – ENE 77 en 78 vormen een geïsoleerde (mini-)kluster;
- 5 – de onder 2 en 3 genoemde klusters worden in een later stadium samengevoegd.

Dit is in overeenstemming met onze schaalanalyse (een iteratieve faktoranalyse volgens Boon van Ostade⁹ leverde een vergelijkbaar resultaat op).

Maken we nu de balans op, dan kunnen we ons wel achter Heunks EA-schaal stellen.

Met betrekking tot de EMA- en EMS-schalen kunnen we ook na weglating van de meest ambivalente (?) items (57, 64 en 69) niet tot een oordeel komen. Omdat ons de voor deze items relevante theoretische achtergrond ontbreekt, is het reëel Heunks t.a.v. deze twee schalen voorlopig te volgen.

De ED-schaal blijft voor ons een raadsel: bij Heunks een duidelijke kluster, bij ons items die niet van EMA (86 en 87) c.q. EMS (88) te onderscheiden zijn.

Hier past een opschorting van ons oordeel.

De bij onze schaalanalyse gevonden lage H-waarde voor de ENE-schaal (.24) lijkt, samen met de exploratieve resultaten van Heunks en onszelf, onze eerdere konklusie te wettigen dat de ENE items niet als schaal kunnen worden opgevat.

9

Daarmede is wat ons betreft, het schaalanalytisch verhaal niet ten einde. Met het niet nagaan van de robuustheid der schalen heeft Heunks verzuimd na te gaan in hoeverre zijn schalen ook voor subgroepen uit de populatie als meetinstrumenten gebruikt kunnen worden. Zijn verweer dat zulks niet is gedaan omdat

Tabel 5:

Overzicht resultaten kluster-analyse op elektorale items van het hoofdonderzoek (de items worden per kluster genoemd in volgorde van opname daarin) (x = niet in een schaal van het hoofdonderzoek opgenomen)

kluster	1	2	3	4	5
	item betr. nr. schaal	item betr. nr. schaal	item betr. nr. schaal	item betr. nr. schaal	item betr. nr. schaal
	93 EA	52 EMA	65 EMS	86 ED	80 ENK
	94 EA	82x ENK	69 EMS	88 ED	81x ENK
	92 EA	71 EMS	57x EMA	77 ENE	78 ENE
	97 EA	63 EMA	74 EMS	87 ED	
		62 EMA	64x EMA	90 ED	

'... al een heel programma van meer algemeen gangbare kontrôles op de schalen was afgewerkt...' (Heunks' kritiek, punt 8)

lijkt ons onjuist omdat die kontrôles wellicht meer algemeen gangbaar zijn (is dat trouwens een kwaliteits-aanduiding?), maar *niet* de robuustheid betreffen.

Interessant is de opmerking uit Heunks repliek (punt 10), dat te laag uitvallende schaalbaarheid pas interessant wordt

'... wanneer, net zoals bij elaboratie van tabellen, een duidelijke specificatie optreedt of spuriousness kan worden vastgesteld, m.a.w. in dit geval de schaalbaarheid geheel (sic) zou ontbreken bij sommige, resp. alle subgroepen.'

Maar om na te gaan of dat het geval is zal men enig onderzoek moeten plegen en dat heeft Heunks nu juist nagelaten.

Daarom is hem ontgaan dat:

- schaal MMA niet robuust is voor respondenten met verschillend opleidingsnivo en ook niet voor verschillende leeftijdskategorieën;
- schaal MN niet robuust is voor leeftijd, opleiding en geslacht;
- schaal EMS en ED niet geheel robuust zijn voor opleidingsnivo.

(Deze resultaten zijn gebaseerd op de chi-kwadraat verdeelde statistiek zoals door Mokken in zijn proefschrift voorgesteld.)

Hoe ernstig dit alles is zal iedere toekomstige gebruiker zelf dienen na te gaan, maar het lijkt ons uitgesloten dat een schaal als bijvoorbeeld MN die met betrekking tot een aantal veel gebruikte indelingen (naar: leeftijd, opleiding, geslacht) absoluut niet robuust is, een grote toekomst tegemoet gaat. Waar Heunks in latere hoofdstukken met behulp van schaalcores hypothesen toetst waarbij indelingen van de respondenten naar bijvoorbeeld opleiding en geslacht een rol spelen, zou een zekere bezorgdheid van zijn kant niet onbegrijpelijk zijn.

10

We komen nu tot de volgende samenvatting:

schaal	commentaar
MMA	zwakke schaal, echter niet robuust
MMS	onvoldoende schaalkwaliteit (I = .24)
MN	zwakke schaal, echter in geen enkele opzicht robuust
MD	onvoldoende schaalkwaliteit (I = .25)
EMA	dimensies
EMS	niet geheel robuust { onbevredigend
ED	niet geheel robuust { te scheiden
ENE	onvoldoende schaalkwaliteit (I = .36) maar: H = .24
ENK	" " (I = .16)
EA	bruikbare, zij het triviale, schaal

11

Omdat Heunks aan het voorgaande minder zwaar tilt dan wij en konkludeert dat de schalen (met uitzondering van de ENK-schaal) als eenduidige schalen kunnen worden beschouwd gaat hij over tot het nagaan van de betrouwbaarheid en de geldigheid van de schalen. Voor de resultaten daarvan laten we zoveel mogelijk Heunks zelf aan het woord.

De betrouwbaarheid van de schalen werd nagegaan met behulp van de 'split-half-methode' (met Spearman-Brown correctie). De split-half-korrelaties bedroegen 'gemiddeld slechts .52 met een minimum van .41.' (Heunks, p. 279)

'We moeten daarom¹⁰ konkluderen dat de betrouwbaarheid van de in het hoofdonderzoek gebruikte schalen veel te wensen overlaat.'

De geldigheid van de schalen: over de *face-validity* en de *content-validity* is Heunks kort. Te kort wellicht indien men bij voorbeeld enige pagina's later leest hoe de EA-schaal over boord wordt gezet. Het aantonen van *concurrent-validity* blijft, door het ontbreken van geschikte gedragsvariabelen, voornamelijk in goede bedoelingen steken: over 8 van de 10 schalen wordt helemaal niets gezegd.

Tenslotte doet Heunks een poging iets over de *construct-validity* te achterhalen. Hij zegt daarover: 'Met name de methode van de 'multitrait-multimethod-matrix' kan op sommige van onze schalen, zij het op onvolmaakte wijze worden toegepast.' (Heunks, p. 280)

Heunks' geldigheids- en betrouwbaarheidsonderzoek geven ons geen aanleiding al onze voorgaande bedenkingen verder maar te vergeten.

12

Op grond van het voorgaande komen wij tot de konklusie dat de tien schalen van Heunks (nog) niet voor een breder gebruik kunnen worden aanbevolen.

Onze reactie 'Tien kleine negertjes' gold Dekkers opmerking: 'Verder zal degene die graag gebruik maakt van attitudinale schalen om de kwaliteit van deze schalen en van de resultaten van deze toepassing ervan beslist zijn voordeel mee kunnen doen.' Onze uitvoerige reactie op Heunks' verhaal heeft mede ten doel duidelijk te maken waarom wij het niet met Dekkers aanbeveling eens zijn. Niet omdat wij ons als een stel door simplisme getroffen schaal-malloten blind staren op een paar I- of H-koëfficiënten. Met Heunks zijn wij van mening dat schaalconstructie een gekompliceerde veelomvattende aangelegenheid is. Het zou daarbij echter van simplisme getuigen dubieuze schaal kwaliteiten, zoals wij die hierboven signaleerden, af te doen als 'het automatisme van een bepaalde rekenprocedure.'

Noten

- 1 Ht = Loevinger's H koëfficiënt voor de gehele schaal.
PPRt = Jackson's PPR voor de gehele schaal.
- 2 De nummers tussen haakjes geven de afzonderlijke items aan zoals die in bijlage I (Heunks, pp. 375-383) staan vermeld.
- 3 Oorspronkelijk zat item nr. 80 ook in de ENK-schaal in plaats van in de ENE-schaal.
- 4 STAP, 1976.
- 5 STAP, 1976.
- 6 Tussen haakjes staat aangegeven de bij elk item behorende a priori schaal van Heunks. De items staan in de volgorde waarin ze in 'schaal' worden opgenomen.
- 7 Wel kan gesteld worden dat het onmogelijk bleek de items uit verschillende schalen te 'scheiden'. Klusteranalyse (Boon, v. Ostade en Johnson) bevestigde dit beeld.
- 8 Overigens is kluster-analyse een dimensielose techniek en als zodanig niet als een validering van m.b.v. schaalanalyse gevonden dimensies te gebruiken.
- 9 STAP, 1976.
- 10 Heunks noemt als redenen: gering aantal items en relatief hoge non-response.

Literatuur

Heunks, F. J., *Aliëntatie en stemgedrag*, 1973.

Mokken, R. J. A., *A theory and procedure of scale analysis*, Den Haag, 1970.

White, B. W. and E. Saltz, 'Measurement of reproducibility', in: G. M. Maranell (ed.), *Scaling, A source book for behavioral Scientists*, Chicago, 1974.

STAP: *Statistical Appendix by SPSS*. Technisch Centrum FSW, Universiteit van Amsterdam, 1976.

Nawoord

door F. J. Heunks

Over het algemeen kan ik het met de inhoud en met name met de teneur van de tweede bijdrage van Boon en Niemöller eens zijn. Zij blijken daarin nu ook oog te hebben voor andere aspecten van schaalontwikkeling dan alleen het determinisme van een schaalbaarheidskoëfficiënt. Overigens heb ik nooit beweerd dat deze 'leeringen van Mokken' simpele zielen zijn, doch slechts dat hun eerste bijdrage in het algemeen leed aan simplisme.

Ten aanzien van twee onderdelen van hun laatste betoog moet zeker nog een opmerking worden gemaakt. Tegenstrijdigheid in mijn standpunt aangaande zogenaamde zoekprocedures is slechts schijnbaar. Ik verwerp 'het selekteren van een kleine subset van items uit een groter aantal items', en bedoel daarmee dat men maat moet houden. Afgezien van vooronderzoekingen, zou men bij voorbeeld als vuistregel kunnen hanteren dat hoogstens één derde deel van een set items die samen een schaal zouden moeten vormen, daar alsnog uit verwijderd mogen worden, bij de vaststelling van de schaalbaarheid in een bepaald onderzoek.

Het tweede detailpunt is opnieuw de robuustheid van een schaal. De waarde daarvan heb ik zeker niet ontkend, maar de wijze van hanteren van dit criterium blijft vraagtekens oproepen. Ten aanzien van welke variabelen wordt robuustheid nagegaan? Wat zijn de grenzen tussen robuust, niet geheel robuust en niet robuust? Afgezien van verdere details waar wederzijds kritiek mogelijk is, heeft de gehele discussie ook tot belangrijke overeenstemming geleid. Ook ik heb mij indertijd wat verbaasd over de 'kwaliteiten' zonder meer van deze schalen, zoals gesteld in de recensie. Mijn opposanten zullen het er mee eens zijn dat daar beter had kunnen worden gewezen op de kwaliteit van de poging om op basis van veel voorafgaande literatuur en empirische studies inzake aliëntatie te komen tot een meer verantwoorde set schalen, hoewel deze mede door een aantal ongunstige omstandigheden in het hoofdonderzoek nog niet, zo ooit, als definitief kunnen worden beschouwd. Ik zou mijn opposanten willen aansporen om op een omzichtige wijze hieraan mee te helpen voortbouwen om tot verdere consolidatie van de betreffende aliëntatieschalen te komen.