

RASCH SCHAAL CONSTRUCTIE MET BEHULP VAN  
EEN PASSINGSINDEX VOOR PERSONEN \*

Herbert Hoijtink  
Vakgroep S&M FSW, RU Groningen  
Oude Boteringestraat 23, 9712 GC Groningen  
050-636199

SUMMARY

A possible cause for the bad fit of a group of items within the unidimensional scaling model of Rasch, could be the presence of persons with a score pattern that is very unlikely to occur under the Rasch model. This paper describes how detection and removal of those unlikely answer patterns can be used in constructing a Rasch scale. The method was applied to three empirical scales and was successful in constructing a Rasch scale for two of them.

INLEIDING

Vanaf 1974 heeft men bij het instituut voor huisartsengeneeskunde te Groningen onderzocht wat aanstaande huisartsen makkelijke en moeilijke verrichtingen vinden. Elke aanstaande huisarts krijgt daartoe een vragenlijst voorgelegd waarop voor elke verrichting op een vierpuntsschaal, variërend van matige tot goed, aangegeven moet worden in hoeverre men deze denkt te beheersen. Omdat alleen het goed beheersen van een verrichting als een positief resultaat van de opleiding gezien wordt, krijgen in een hercodering de goed beheerste verrichtingen de score 1, de andere krijgen de score 0.

Na deze dichotomisering kan per verrichtingsgroep (bijvoorbeeld neurologische, gynaecologische en oogheeskundige verrichtingen) een Rasch analyse worden uitgevoerd. Daarbij wordt onderzocht of de verrichtingen in een groep eenzelfde soort vaardigheid veronderstellen en hoeveel van die vaardigheid nodig is om een bepaalde verrichting met een goede kans op succes uit te kunnen voeren (zie Wright & Stone, 1979 voor een uitgebreide beschrijving van het Rasch model).

De resultaten zijn nogal wisselend, sommige modeltoetsen geven aan dat de verrichtingen een Rasch schaal vormen (er is voor alle verrichtingen in een groep eenzelfde soort vaardigheid nodig en er kunnen makkelijke en moeilijke verrichtingen onderscheiden worden), andere geven aan dat de verrichtingen geen schaal vormen (Groenier, 1981). Gezien deze tegenstrijdige resultaten en het invulgedrag van sommige aanstaande huisartsen, vermoedt Groenier dat zijn data "vervuild" zijn door afwijkende personen, d.w.z. personen die de vragenlijst niet al te serieus ingevuld hebben. Het zou interessant zijn te kijken hoe goed het Rasch model past nadat deze personen uit de datamatrix verwijderd zijn.

Aangezien er destijds nog niet zoveel bekend was over passingsindexen voor personen, is dit achterwege gelaten. Momenteel is deze kennis er wel, van een aantal indexen zijn verdeling en robuustheid geëvalueerd (Hoijtink, 1986) en naar aanleiding van dit onderzoek is er een nieuwe

\* Met dank aan Drs. K.H. Groenier, Prof. Dr. I.W. Molenaar en Dr. W.H. van Schuur voor hun hulp bij de totstandkoming van dit artikel.

passingsindex ontwikkeld (Molenaar & Hoijtink, 1987).

In dit paper wordt beschreven hoe met behulp van deze nieuwe passingsindex geprobeerd is afwijkende personen op te sporen om ze daarna uit de analyse te verwijderen, waarna er misschien wel een schaal voor de gynaecologische, neurologische en oogheelkundige verrichtingen gevonden wordt.

#### SCHAALCONSTRUCTIE EN MODELTOETSEN

De beginsituatie bij schaalconstructie volgens het model van Rasch ziet er altijd als volgt uit :

- I men heeft een groep items die verondersteld worden een bepaalde latente trek te definiëren (b.v. moeilijkheid van gynaecologische verrichtingen),
- II men heeft een groep respondenten waarvan aangenomen wordt dat ze een bepaalde hoeveelheid van deze latente trek bezitten (b.v. verwachte vaardigheid op gynaecologisch gebied).

Vervolgens wordt meestal alleen getoetst :

- of alle items samen een schaal vormen,
  - of elk afzonderlijk item voldoet aan de eisen van het model van Rasch.
- Een voor een worden nu steeds de slechtst passende items verwijderd, totdat elk item apart en alle items samen, passen binnen het model van Rasch.

Nadeel van deze methode is dat buiten beschouwing wordt gelaten dat de items die verwijderd worden, wel eens slecht zouden kunnen passen omdat er enkele personen zijn die een negatieve invloed op het passen van deze items hebben. Deze personen hebben geen gebruik gemaakt van bijvoorbeeld de verwachte vaardigheid op gynaecologisch gebied bij het beantwoorden van de vragenlijst. Verwijdering van deze personen uit de datamatrix leidt tot betere schattingen van de item- en persoonsparameters, en verbetert het passen van het Rasch model.

Een en ander lijkt misschien op het aanpassen van de data aan het model, maar er zijn tal van situaties waarin het verwijderen van personen uit de datamatrix te rechtvaardigen valt. Als 80 % van de respondenten een vragenlijst serieus invuld en 20 % heeft er eigenlijk geen zin in en doet het at random, dan kunnen die 20 % er de oorzaak van zijn dat het Rasch model niet past. Als deze 20 % uit de datamatrix verwijderd worden, kunnen op grond van de overblijvende 80 % goede schattingen van item- en persoonsparameters en het passen van het model gemaakt worden. Als daarna in test situaties de vragenlijst opnieuw afgenomen wordt, kan voor elke geteste persoon nagegaan worden of hij bij de 20 % random invullers hoort, en het model en de daarbijbehorende vaardigheidsschattingen op hem dus niet van toepassing zijn, of dat hij bij de 80 % serieuze invullers hoort.

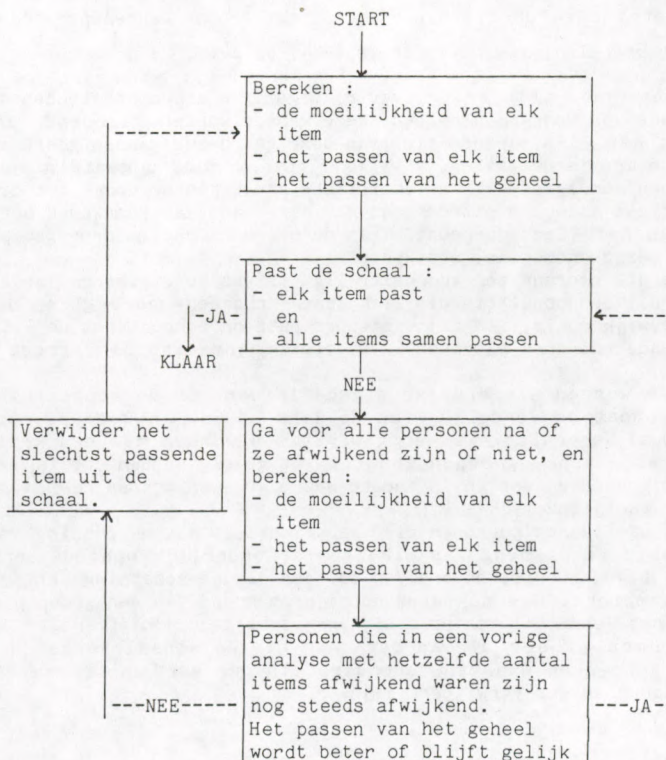
Een andere hierop lijkende situatie ontstaat als de steekproef voornamelijk bestaat uit Nederlanders, maar er ook immigranten en gastarbeiders in voorkomen. Op grond van taalproblemen of culturele verschillen kan de moeilijkheid van de items voor deze twee groepen totaal anders liggen, het lijkt dan ook redelijk om de steekproef te splitsen, waarna binnen elke groep de modelparameters geschat kunnen worden.

Statistisch gezien zullen de antwoordpatronen van de 20 % en de immigranten en de buitenlanders zich kenmerken door een lage kans van voorkomen gegeven de modelparameters zoals geschat over de 100 % of gegeven de totale steekproef.

In de rest van dit hoofdstuk zal een methode beschreven worden die afwijkende personen uit de datamatrix verwijderd, en zo leidt tot een beter passen van het Rasch model en betere schattingen van item- en persoonsparameters.

In schema I is een overzicht gegeven hoe Rasch schaal constructie met behulp van een passingsindex per persoon in zijn werk gaat. Eerst worden voor elk item de moeilijkheid berekend. Daarna wordt het passen getoetst met de binomiale toets (Molenaar, 1983). Tevens wordt getoetst of alle items tezamen passen, hiervoor wordt de Andersen toets gebruikt (twee groepen, hoge versus lage score, Gustafsson, 1979). Voor de berekeningen is gebruik gemaakt van het programma PML (Gustafsson, 1979; Molenaar 1983). In principe kunnen ook andere en meer toetsen gebruikt worden, dat maakt voor de idee achter het verwijderen van afwijkende personen niet uit. Een goede aanvulling, helaas in PML niet aanwezig, zou de Q2-toets waarmee de eendimensionaliteit van de items getoetst kan worden zijn (Van den Wollenberg, 1982).

SCHEMA I : Rasch schaal constructie met behulp van een passingsindex per persoon.



Hierna wordt de schaal geevalueerd. Geeft de Andersen toets aan dat alle items tezamen passen (P-waarde groter dan .05, of als men dit te streng vindt .01) en passen alle items afzonderlijk (gestandaardiseerde binomiale toetsingsgrootte tussen -1.96 en 1.96 bij toetsing met  $\alpha = .05$  of tussen -2.56 en 2.56 bij toetsing met  $\alpha = .01$ ), dan is men klaar met de analyse en is er een schaal gevonden.

Past het geheel niet of past een afzonderlijk item niet, dan neemt het constructieverloop een vervolg, voor elke persoon wordt onderzocht of zijn antwoordpatroon afwijkend is. Daartoe wordt gegeven de itemmoeilijkheden een steekproef gesimuleerd die Rasch homogeen is en uit 2000 antwoordpatronen per score groep bestaat. Voor elk van deze patronen wordt de kans gegeven de ruwe score berekend :

$$P(X|r) = \epsilon_1^{x_1} \epsilon_2^{x_2} \dots \epsilon_k^{x_k} / \gamma_r(\epsilon) \quad (1),$$

$X = (x_1, x_2, \dots, x_k) =$  antwoordpatroon,

$r = \sum_{i=1}^k x_i =$  ruwe score,

$\epsilon_k = \text{EXP}(-\sigma_k) =$  multiplicatieve itemmoeilijkheid,

$\gamma_r(\epsilon) =$  symmetrische functie die ervoor zorgt dat de kansen per scoregroep tot 1 optellen, zie b.v. Fisher, 1974, pp 241).

Per scoregroep zijn er (n over r) mogelijke antwoordpatronen die elk een zekere kans van voorkomen hebben. Deze kansen worden geordend van hoog naar laag en het 95-ste percentiel van deze geordende kansen geeft in elke score groep de kritische waarde. Afwijkend zijn wordt nu gedefinieerd als het binnen een scoregroep uit de empirische steekproef behoren tot die personen waarvan het antwoord patroon een kleinere kans van voorkomen heeft dan de waarde van het 95-ste percentiel in de overeenkomstige scoregroep in de gesimuleerde (Rasch homogene) steekproef.

De idee die hieraan ten grondslag ligt is dat de empirische steekproef betrokken is uit een populatie die zich Rasch homogeen gedraagt en dat een persoon dus afwijkend is, als zijn antwoord patroon behoort tot de 5 % minst vaak voorkomende antwoord patronen in de overeenkomstige scoregroep in de populatie.

Personen worden als afwijkend gedefinieerd in de context van hun scoregroep, omdat de verdeling van de kans op een antwoordpatroon per scoregroep nogal verschilt. Als er gewerkt zou worden met een kritische waarde voor alle score groepen zou dit tot gevolg hebben dat in bepaalde score groepen mensen zelden en in andere score groepen mensen heel vaak afwijkend zijn (Hoijtink, 1986).

Nadat de afwijkende personen zijn verwijderd, wordt de passing van het geheel, evenals de passing van elk item afzonderlijk opnieuw berekend. Uiteraard leidt een hernieuwde analyse tot een nieuwe schatting van item en persoons parameters, die nu berekend zijn op grond van een groep personen waaruit tenminste een deel van de afwijkers verwijderd is. Hierdoor hangen de berekeningen minder af van personen die de schaal verkeerd geïnterpreteerd hebben en waarvan verondersteld mag worden dat ze tot de afwijkers behoren en dus verwijderd zijn.

Daarna wordt gekeken of de data na het verwijderen van afwijkende personen nog consistent zijn. Op twee manieren kunnen inconsistenties optreden :

- Worden met hetzelfde aantal items meerdere malen afwijkende personen verwijderd (in de tweede verwijderonde wordt het afwijken berekend op grond van de itemmoeilijkheden geschat op grond van de antwoorden van de personen die in de eerste verwijderonde niet afwijkend zijn bevonden) dan moeten personen die in de eerste ronde afwijkend waren, dat in de tweede ronde ook zijn.
- Zijn personen afwijkend op grond van de "slechtere schattingen" van de itemmoeilijkheden in de eerste ronde, dan moeten ze dat ook zijn op grond van de herberekende "betere" schattingen, iemand die afwijkend is moet dat dus blijven, is dit niet meer het geval dan moet er opgehouden worden met het verwijderen van personen omdat de onderliggende structuur in de data dan geweld aan wordt gedaan en het waarschijnlijker is dat er slecht passende items in plaats van nog meer slecht passende personen tot de datamatrix behoren.
- Door het verwijderen van afwijkende personen, moet een schaal als geheel beter gaan passen. Als de toets op het passen van de schaal als geheel na het verwijderen van afwijkende personen aangeeft dat de schaal slechter is gaan passen, is het raadzaam op te houden met het verwijderen van personen.

Wordt aan bovenstaande twee voorwaarden voldaan, dan moet de kwaliteit van de schaal opnieuw getest worden. Wordt aan deze voorwaarden niet voldaan, dan ligt het niet passen van de schaal niet alleen aan de aanwezigheid van afwijkende personen, maar vermoedelijk ook aan de aanwezigheid van afwijkende items. De volgende stap is dan het slechtst passende item uit de analyse te verwijderen.

Voor het verwijderen of behouden van items zijn een aantal vuistregels te geven :

- Als afwijkende personen verwijderd worden, en dat heeft tot gevolg dat een item slechter gaat passen, dat was een deel van het passen van dat item gebaseerd op de aanwezigheid van die afwijkende personen. Op grond daarvan kan overwogen worden dat item uit de analyse te verwijderen.
- Items die slecht passen, ook na de verwijdering van afwijkende personen, voldoen waarschijnlijk niet aan de eisen van het schaalmodel.
- Items die beter gaan passen na de verwijdering van afwijkende personen konden wel eens slecht passen door die afwijkende personen. Deze items moeten in eerste instantie niet verwijderd worden.

Tot zover de bespreking van het schema. In het volgende hoofdstuk zal uiteen worden gezet wat de eerste ervaringen zijn met deze methode van schaalconstructie.

## DE SCHAALCONSTRUCTIE

In dit hoofdstuk zal beschreven worden hoe de schaalconstructie voor de oogheelkundige, neurologische en gynaecologische verrichtingen verliep.

### De oogheelkundige vaardigheidsschaal

De analyse begon met 23 items betreffende neurologische vaardigheden. De aanstaande huisartsen moesten aangeven hoe goed ze elke vaardigheid beheersten zoals daar waren ; het omklappen van de oogleden, oogspiegelen, oogspoelen en verwijderen van corpus alienum.

Deze items vormden in eerste instantie geen schaal ; er zijn items met te hoge en te lage waarden op de gestandaardiseerde binomiale toetsingsgroottheid ( $< -1.96$  en  $> 1.96$ ) en de overschrijdingskans van de Andersen test wijst op niet passen (kolom 2 TABEL 1). Kolom 3 geeft de resultaten nadat afwijkende personen verwijderd zijn. Van de items die slecht pasten ; 1, 3, 7, 8 en 10 gaan 1, 3, 7 en 10 beter passen. Op grond van de overgebleven personen (455) werden de itemmoeilijkheden opnieuw geschat, waarna opnieuw voor alle 492 personen getest werd of ze afwijkend zijn of niet.

Dit leidde tot een groter aantal afwijkenden dan in eerste instantie (439 niet afwijkende antwoordpatronen), waarbij alle personen die in eerste instantie afwijkend waren bevonden, ook nu als afwijkend werden geclassificeerd. In kolom 4 is te zien dat de Andersen test aangeeft dat de schaal als geheel slechter gaat passen. Er wordt dus opgehouden met het verwijderen van personen en het slechtste item, 8, wordt uit de schaal verwijderd.

TABEL 1 : De oogheelkunde schaal.

Voor verschillende analyses de gestandaardiseerde binomiale waarden, de overschrijdingskans van de Andersen toets (hoge versus lage scores) en aantal personen in de analyse:

Notatie : I23P0 23 items, 0 maal personen verwijderd. I22P1 22 items, 1 maal personen verwijderd.

Itemnr	Kolom					
	1	2	3	4	5	6
	I23P0	I23P1	I23P2	I22P0	I22P1	
1	4.38	2.01	1.61	3.90	1.21	
2	1.55	1.19	0.57	1.68	0.50	
3	2.55	1.94	2.42	2.34	2.01	
4	0.14	0.39	-0.41	0.30	0.73	
5	0.38	1.26	1.45	0.23	0.89	
6	0.12	0.37	0.30	-0.14	0.07	
7	-2.07	-1.54	-1.26	-1.38	-0.86	
8	-2.62	-2.43	-2.40			
9	0.74	0.39	0.59	0.28	-0.17	
10	-2.36	-2.12	-1.78	-2.32	-1.77	
11	-0.80	-0.74	-0.56	-0.77	-0.59	
12	1.00	0.66	0.69	0.99	0.47	
13	-0.09	0.22	-0.05	-0.56	-0.09	
14	0.12	0.21	0.37	-0.05	0.23	
15	-0.61	-0.93	-0.81	-0.72	-0.70	
16	-0.87	-0.99	-0.97	-0.94	-1.09	
17	-0.46	-0.89	-0.72	-0.86	-1.06	
18	-0.31	-0.05	-0.18	-0.69	-0.67	
19	1.91	1.05	1.26	1.54	1.05	
20	-1.39	-1.20	-1.48	-1.49	-1.51	
21	1.16	1.20	1.22	1.16	1.20	
22	-1.12	-0.70	-1.14	-1.17	-1.07	
23	-0.33	-0.48	-0.72	-0.31	-0.67	
Andersen	0.00	0.30	0.29	0.01	0.37	
Aantal	492	455	439	492	453	

De resultaten voor de resterende 22 items en 492 niet nul- of perfect scorenden, staan in kolom 5. Item 1 en 3 passen slecht en de totale schaalbaarheid is ook nog onder de maat. Kolom 6 geeft de resultaten na eenmalige verwijdering van afwijkende personen ; de totale schaalbaarheid is goed. Alleen item 3 past marginaal, bij toetsing met  $\alpha = .01$ , is een waarde van 2.01 op de gestandaardiseerde binomiale toetsingsgrootte echter acceptabel, waarna besloten is dit item niet uit de schaal te verwijderen. Herhaalde verwijdering van afwijkende personen leidde tot een verslechtering van de totale schaalbaarheid, waarna besloten werd het hierbij te laten. Van de 492 personen werden er voor de 22 item schaal 39 (8 %) uit de analyse verwijderd. Onder het Rasch model zouden 5 % afwijkenden normaal geweest zijn. Meest opvallend is de verbetering in fit van item 1.

De neurologische vaardigheidsschaal

In het begin waren er 18 items betreffende neurologische vaardigheden zoals ; het testen van de hersenzenuwen, beoordelen van het bewustzijn en beoordelen van de motoriek.

TABEL 2 : De neurologie schaal.

Voor verschillende analyses de gestandaardiseerde binomiale waarden, de overschrijdingskans van de Andersen toets (hoge versus lage scores) en aantal personen in de analyse.

Notatie : I18P0 18 items, 0 maal personen verwijderd. I16P2 16 items, 2 maal personen verwijderd.

Kolom										
	1	2	3	4	5	6	7	8	9	10
Item	I18P0	I18P1	I17P0	I17P1	I17P2	I17P3		I16P0	I16P1	I16P2
1	-1.04	-1.26		-1.12	-1.19	-0.87	-1.08	-1.00	-1.39	-0.98
2	0.00	0.08		0.25	0.24	0.51	0.62	0.24	1.18	0.73
3	0.57	-0.73		0.65	-0.09	-0.20	0.02	0.79	0.22	0.13
4	-0.46	-0.28		0.73	0.24	-0.55	-0.27	0.00	0.25	0.55
5	0.03	-0.18		1.20	0.34	-0.29	-0.12	0.48	0.37	0.58
6	-0.45	-0.58		-0.93	-0.70	-0.46	-0.13	-0.11	0.18	-0.54
7	-0.26	-0.35		-0.58	-0.24	-0.05	0.10	-0.71	0.18	0.06
8	-2.17	-1.91		-1.86	-1.81	-1.31	-0.92	-2.15	-1.65	-1.19
9	-0.28	0.26		0.18	0.24	-0.61	-0.12	0.70	1.42	-0.24
10	0.59	1.13		1.03	1.95	1.29	1.67	1.23	1.83	0.93
11	-1.51	-1.44		-1.62	-1.20	-0.55	-0.90	-0.66	-1.03	-0.16
12	-2.04	-2.25		-2.66	-2.59	-2.35	-3.12	-1.44	-1.12	-1.60
13	-2.07	-2.37		-2.47	-2.37	-2.17	-2.45	-1.65	-1.44	-1.32
14	-0.66	0.17		-0.10	0.29	-0.08	0.18	-0.01	-1.82	-0.82
15	3.76	4.94								
16	2.58	1.34		3.30	2.67	3.30	2.96			
17	0.14	0.20		0.11	0.07	-0.38	-0.17	0.42	-0.48	-0.29
18	2.00	1.07		4.09	1.29	0.09	-0.10	3.65	2.13	0.40
Andersen										
	0.00	0.00		0.00	0.02	0.45	0.00	0.00	0.01	0.92
aantal	405	371		396	363	332	308	390	348	318

In kolom 4 van tabel 2 staat de situatie nadat item 15 verwijderd is. Items 12, 13, 16 en 18 passen slecht. Na herhaald verwijderen van afwijkende personen, waarmee gestopt werd toen de p-waarde van de Andersen test daalde ( $0.45 > 0.00$ ), vertoont de schaalbaarheid van items 12, 13 en 18 een positieve trend, item 16 blijft slecht passen en wordt verwijderd. Er resteert een schaal van 16 items die na tweemaalige verwijdering van afwijkende personen goed blijkt te passen.

TABEL 3 : De gynaecologie schaal.

Voor verschillende analyses de gestandaardiseerde binomiale waarden, de overschrijdingskans van de Andersen toets (hoge versus lage scores) en aantal personen in de analyse.

Notatie : I32P0 32 items, 0 maal personen verwijderd. I26P1 26 items, 1 maal personen verwijderd.

Kolom							
1	2	3	4	5	6	7	
Item	I32P0	I32P1	I31P0	I31P1	I26P0	I26P1	
1	-0.43	-0.18	-0.39	-0.01	-0.53	-0.60	
2	0.58	0.27	0.72	0.44	0.46	0.38	
3	1.33	1.17	1.46	1.25	0.89	1.02	
4	-1.65	-2.33	-1.45	-2.12			
5	1.49	1.87	1.28	1.50	1.42	1.91	
6	0.79	0.16	0.68	0.68	0.66	0.02	
7	-1.65	-2.33	-1.45	-2.12			
8	-1.85	-1.73	-1.69	-1.47	-1.55	-1.52	
9	1.64	0.55	1.90	0.55	2.76	0.29	
10	-0.89	-0.36	-0.67	-0.27	-0.49	-0.11	
11	-1.04	-0.87	-0.72	-0.61	-1.30	-1.15	
12	-0.63	-0.20	-0.59	-0.31	-0.49	0.04	
13	-2.37	-2.12	-2.18	-1.99	-2.04	-1.83	
14	2.52	2.00	2.89	2.38	3.69	2.39	
15	-1.06	-0.32	-1.10	-0.40	-0.80	-0.16	
16	0.33	0.38	0.41	0.36	0.29	0.22	
17	-1.41	-1.19	-1.18	-0.90	-1.45	-1.35	
18	0.76	1.67	0.27	0.87	0.04	0.01	
19	-3.15	-2.75	-3.20	-2.64	-2.64	-2.19	
20	-1.02	-0.69	-0.84	-0.74	-0.95	-0.76	
21	-2.41	-2.64	-2.25	-2.33			
22	-0.17	-0.37	-0.11	-0.38	-0.48	-0.64	
23	-0.99	-0.52	-1.32	-1.13	-1.99	-1.66	
24	4.72	5.39					
25	4.07	3.21	5.33	4.60			
26	1.28	2.02	1.55	1.69			
27	1.72	1.36	1.84	1.52	2.39	2.02	
28	-1.94	-1.83	-1.41	-1.29	-0.94	-0.69	
29	-0.32	-0.21	-0.19	0.07	-0.29	-0.16	
30	-0.63	-0.55	-0.42	-0.53	-0.62	-0.67	
31	-0.13	-0.09	-0.29	-0.27	0.14	0.20	
32	-0.13	-0.36	-0.19	-0.37	0.06	-0.31	
Andersen	0.00		0.00				
Aantal	481	446	479	447	476	448	



Van de 390 personen die op de 16 item grote schaal niet nul- of perfect scoorden bleven er 318 over, 72 personen (18 %), zijn uit de schaal verwijderd, er zaten dus 13 % meer afwijkende personen in de data dan modelmatig verwacht werd.

#### De gynaecologische vaardigheidsschaal

Voorbeelden van de vaardigheden die in de 32 items werden onderzocht zijn ; vaginaal toucher, breken van de vliezen, ruptuur hechten en het begeleiden van de uitdrijving bij stuitligging.

Van de gehele analyse is hier slechts een deel weergegeven in Tabel 3. Na verwijdering van afwijkende personen verslechterde het passen van veel items. Er werd een item verwijderd waarna het zelfde beeld weer terugkeerde. Dit proces zette zich voort tot er nog maar 22 items over waren waarna de analyse gestopt is. De niet-schaalbaarheid van de gynaecologische verrichtingen ligt blijkbaar niet aan de aanwezigheid van afwijkende personen. Er zal naar een andere oorzaak gezocht moeten worden, bijvoorbeeld item karakteristieke curves die zich niet model conform gedragen, of meerdimensionaliteit van de items.

#### DISCUSSIE

Het blijkt mogelijk afwijkende personen op te sporen, hen uit de datamatrix te verwijderen en zo de aan de data onderliggende structuur te vinden. Niet alle slecht passende schalen worden echter "vervuild" door de aanwezigheid van afwijkende personen, zoals blijkt uit de gynaecologische schaal. Wordt er echter vermoed dat deze "vervuiling" wel aanwezig is dan kan deze methode ook in andere toepassingen, soelaas bieden.

Aangezien er niet voldoende proefpersonen aanwezig waren is er in deze studie geen kruisvalidatie van de gevonden schalen uitgevoerd. Dit is echter wel gewenst, het passen van personen en items wordt zo vaak getoetst dat er zeker sprake is van kanskapitalisatie.

Tevens kan het inzichtelijk zijn het al dan niet afwijken te vergelijken met enige achtergrondkenmerken van de respondenten. Er kan dan gezocht worden naar gemeenschappelijke kenmerken van de afwijkers. Echter naar dergelijke gegevens was in dit onderzoek niet gevraagd. Een nadeel hiervan is dat " het statistisch afwijken ", niet inhoudelijk onderbouwd kan worden.

De uiteindelijke conclusie lijkt te zijn dat Rasch schaal constructie met een passingsindex per persoon een goede hulp kan zijn als de data "vervuild" zijn door afwijkende personen\* .

#### LITERATUUR

- Fischer G.H. (1974) Einführung in die theorie psychologischer tests, Bern : Huber.  
 Groenier K.H. (1981) Het meten van technische vaardigheden van arts assistenten, Instituut voor huisartsgeneeskunde Groningen.

\* Geïnteresseerden kunnen het FORTRAN 77 programma voor het berekenen van het passen van de personen, aanvragen bij de auteur.

- Gustafsson J.E. (1979) PML a computer program for conditional estimation and testing in the Rasch model for dichotomous items, Reports from the institute of education, University of Goteburg, no 63.
- Hoijtink H. (1986) Detecting aberrant response patterns in the unidimensional scaling model of Rasch, Heymans Bulletin Psychologische Instituten R.U. Groningen, HB-86-792-SW.
- Molenaar I.W. (1981) Programmabeschrijving van PML (versie 3.1) voor het Rasch-model, Heymans Bulletin Psychologische Instituten R.U. Groningen, HB-81-538-RP.
- Molenaar I.W. (1983) Some improved diagnostics for failure of the Rasch model, Psychometrika, 48, 49-72.
- Molenaar I.W., Hoijtink H. (1987) The many null distributions of person fit indices, Heymans Bulletins Psychologische Instituten R.U. Groningen, HB-87-846-EX.
- Van den Wollenberg A.L. (1982) Two new test statistics for the Rasch model, Psychometrika, 47, 123-140.
- Wright B.D., Stone M.H. (1979) Best test design, MESA PRESS University of Chicago.

Ontvangen: 04-02-1987  
Geaccepteerd: 09-07-1987