

9. Het meten en analyseren met modellen in sociologisch onderzoek

F.N. Stokman

9.1. INLEIDING

In de methodologie van het sociologisch onderzoek doet zich een belangrijke ontwikkeling voor, waardoor het meten van sociale verschijnselen en de analyse van hun onderlinge structuur een veel minder op intuïtie gebaseerd karakter krijgt, dan vroeger het geval was. Dit is te danken aan de ontwikkeling en toepassing van een groot aantal nieuwe meet- en analysemodellen, en een wetenschapstheoretische fundering van het meten in de sociale wetenschappen. Hierbij wordt het meten gezien als het afbeelden van een empirisch relationeel systeem in een numeriek relationeel systeem. Het empirische relationele systeem bestaat uit de objecten, waarin wij geïnteresseerd zijn, en empirische relaties die wij op basis van bepaalde kenmerken tussen die objecten kunnen aanbrengen. In de sociale wetenschappen zijn de objecten vaak personen, maar dat is lang niet altijd het geval. Ons object van meting kan net zo goed bestaan uit sociale organisaties of andere sociale verschijnselen. Als wij in de sociale wetenschappen objecten met elkaar vergelijken op basis van een bepaald kenmerk, dan komen wij zelden verder dan zulke zwakke empirische relaties als 'meer dan' of 'evenveel als' of 'deze twee lijken meer op elkaar dan die andere twee' of 'deze twee lijken net zo veel op elkaar als die andere twee'. De kern van het meetprobleem bestaat nu hieruit, dat wij aan elk object een getal willen toekennen op een zodanige wijze dat bij waarneming van een der bovengenoemde empirische relaties ook de daarmee corresponderende numerieke relatie tussen de toegekende getallen geldt. Wij zullen straks een voorbeeld uitwerken waarin de studietijd in de propedeuse sociologie te Groningen wordt gemeten. Dit is gebeurd door studenten voor elk tweetal studie-onderdelen te vragen aan welke van de twee onderdelen zij meer tijd hebben besteed. Studie-onderdelen waaraan meer tijd is besteed, zullen een hoger getal toegekend moeten krijgen dan studie-onderdelen waaraan minder tijd is besteed. Nu reeds is duidelijk dat dit geen eenvoudige zaak is. Voor één student kan dit al problemen opleveren, wanneer hij 'inconsistent' heeft geantwoord door te zeggen dat hij aan studie-onderdeel A meer tijd heeft besteed dan aan B, aan B meer dan aan C, maar vervolgens stelt dat hij aan C meer tijd heeft besteed dan aan A. Over alle studenten tezamen

wordt het nog moeilijker, want dan moeten alle studenten niet alleen consistent zijn geweest, maar ook overwegend dezelfde ordening van studieonderdelen hebben aangebracht, ongeacht verschillen in vooropleiding en bekwaamheden. Pas als aan deze veronderstellingen is voldaan, is het mogelijk de studie-onderdelen (onze objecten) een rangnummer toe te kennen op basis van de bestede studietijd (kenmerk) op een zodanige wijze dat studenten aan een studie-onderdeel met een hoger toegekend getal (meestal) meer tijd besteedden dan aan een studie-onderdeel met een lager rangnummer. Meer dan een rangorde hebben we echter ook dan nog niet, terwijl wij eigenlijk een getallentoeckenning zouden willen hebben waaruit de feitelijke studietijd af te leiden is. Dit laatste en een toetsing van de veronderstellingen is pas mogelijk door een *theorie* te formuleren, waarin gespecificeerd wordt hoe de antwoorden van de studenten gerelateerd zijn aan de feitelijke, ons onbekende, studietijden. Die theorie moet dan wel zodanig zijn dat wij daaruit metingen voor de studie-onderdelen kunnen afleiden, alsmede restricties waaraan de gegeven antwoorden moeten voldoen en op basis waarvan wij kunnen toetsen of onze theorie wel opgaat. Wordt onze theorie weersproken, doordat de antwoorden van de studenten niet overeenkomstig de theorie gegeven worden, dan hebben wij een onjuiste theorie en dus ook geen metingen. Wordt onze theorie niet weersproken, dan hebben wij zowel een theorie als metingen. Onze *metingen* zijn dus een *toegift van de theorie*. Dergelijke theorieën worden meestal aangeduid met de term *meetmodellen*. Het bovenstaande samenvattend in meer algemene terminologie, bezit een meetmodel de volgende kenmerken:

- het specificeert hoe empirische relaties, ook wel *manifest* gedrag genoemd, gerelateerd zijn aan *latente* kenmerken (in ons voorbeeld hoe de antwoorden van de studenten op de paarsgewijze vergelijkingen van de studie-onderdelen gerelateerd zijn aan de feitelijke, ons onbekende, studietijden). Dit houdt tevens in dat het model de aard van het latente kenmerk specificeert;
- het omvat een *schattingstheorie*, op basis waarvan de ons onbekende grootheden uit de antwoorden kunnen worden geschat. Deze onbekende grootheden worden de parameters van het model genoemd. In ons voorbeeld zijn dat de getallen die aan de studie-onderdelen moeten worden toegekend op basis van de daaraan bestede studietijd;
- het omvat een *toetsingstheorie*, op basis waarvan het model getoetst kan worden aan de hand van de restricties, die uit het model voor het manifeste gedrag zijn af te leiden.

Meetmodellen maken hiermee metingen in de sociale wetenschappen tot een integraal onderdeel van theorievorming. Metingen verliezen daarmee

hun i
meet
den e
(der) v
dijetij
welk
Mo
de soc
sen sc
kenm
ceerd,
van se
de str
model
sultate
doure
van he
gaan c
en str
door J
REL (v
wel een
te toet
(Jöres
bereik

In p
dankzi
graaf S
dellen
men n

9.2. BE

In een
pedeus
ges (19
waarbi
welk v
tot een

hun intuïtieve en willekeurige karakter. Daarenboven stelt een aantal meetmodellen ons in staat tot een hoger meetniveau te komen ook al worden empirische relaties slechts dichotoom (ja, nee) of ordinaal (meer-minder) waargenomen. Zo zullen wij zien dat wij in staat zijn de feitelijke studietijden te schatten, ook al hebben de studenten slechts aangegeven aan welk onderdeel zij meer of minder tijd besteed hebben.

Modellen spelen niet alleen in toenemende mate een rol bij het meten in de sociale wetenschappen, maar ook bij de analyse van de samenhang tussen sociale verschijnselen. Worden in meetmodellen aard van het latente kenmerk en relatie tussen manifest gedrag en latente kenmerk gespecificeerd, in *structurele modellen* specificeert de onderzoeker aard en richting van samenhang tussen verschillende (latente) kenmerken en legt daarmee de *structuur* tussen die verschijnselen vast. Het zijn juist deze structurele modellen die veelal een verbinding leggen tussen empirische onderzoeksresultaten en ruimere theoretische kaders binnen de sociologie (zie bijv. Gadourek, 1982). Ook hier is een schattingstheorie nodig om de parameters van het model te kunnen schatten en een toetsingstheorie om te kunnen nagaan of c.q. hoe goed het model past. Het onderscheid tussen meetmodel en structureel model wordt zeer expliciet gehanteerd in een methode, die door Jöreskog in 1970 is gepubliceerd, en bekend staat onder de naam LISREL (Lineair Structural RELations). Deze methode stelt ons in staat zowel een meet- als een structureel model te specificeren en beide gelijktijdig te toetsen op basis van een meest aannemelijke schatting van de parameters (Jöreskog, 1970; 1977). Hierdoor wordt een volledige modellencombinatie bereikt (zie § 1.4.1).

In paragraaf 9.2 zullen wij eerst een voorbeeld geven van metingen die dankzij inschakeling van een meetmodel tot stand zijn gekomen. In paragraaf 9.3 vatten wij het gehele proces van het meten en analyseren met modellen nog eens samen en expliciteren wij de beslissingen die daarbij genomen moeten worden.

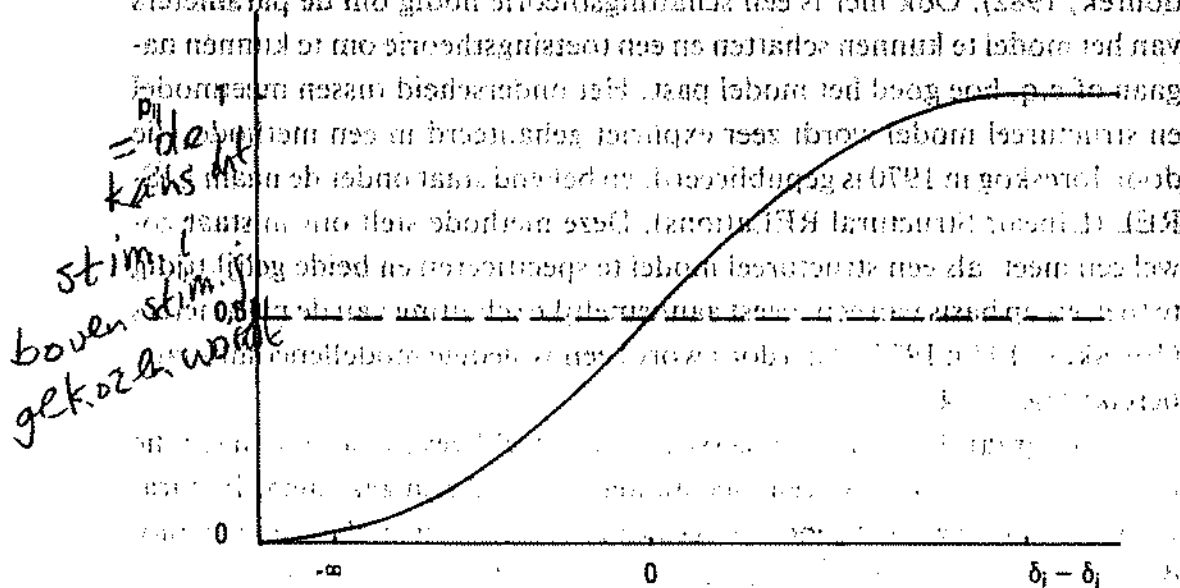
9.2: EEN VOORBEELD

In een onderzoek naar de studiebelasting van studie-onderdelen in het propedeuse jaar 1977/78 bij de sociologie-opleiding te Groningen besloot Rutgers (1978) gebruik te maken van de methode van paarsgewijze vergelijking, waarbij studenten voor elk tweetal studie-onderdelen moesten aangeven, welk van de twee meer studietijd had gevergd. Om vanuit deze gegevens tot een schatting van studietijden te komen, paste hij het meetmodel van

Thurstone (1927) toe en ijkte vervolgens de daaruit resulterend schaalwaarden aan een aantal studie-onderdelen waarvan het aantal werkuren door de studenten direct moest worden geschat.

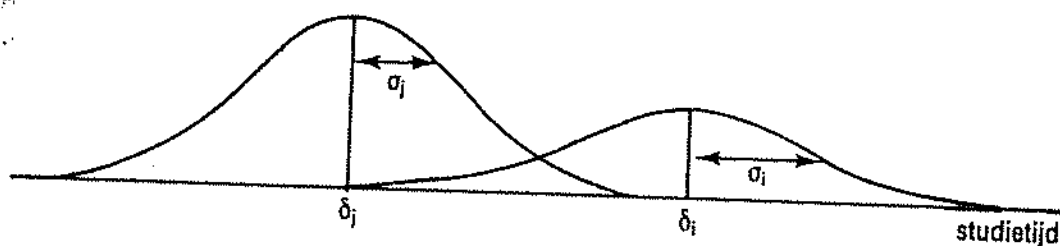
Het model van Thurstone ziet er als volgt uit: Thurstone gaat uit van een ééndimensionaal continuüm waarop elk der stimuli een bepaalde, ons onbekende, waarde bezit. In onze toepassing is het continuüm de studietijd en de stimuli zijn de studie-onderdelen. De waarde van stimulus i op het continuüm geven wij aan met δ_i . Wanneer nu iemand gevraagd wordt voor een tweetal stimuli i en j aan te geven welke van de twee een hogere schaalwaarde heeft, dan zal de kans p_{ij} dat stimulus i boven j gekozen wordt monotoon stijgen met het verschil $\delta_i - \delta_j$. Dit is in *figuur 9.1* aangegeven.

Figuur 9.1: De relatie tussen de kans dat stimulus i boven stimulus j wordt verkozen en het verschil in schaalwaarden $\delta_i - \delta_j$.



Het is redelijk aan te nemen dat de kans p_{ij} gelijk is aan 0,5 als δ_i en δ_j dezelfde waarde op het continuüm hebben. In onze toepassing vergen beide studie-onderdelen dan evenveel studietijd. Is studie-onderdeel i zwaarder dan studie-onderdeel j , dus $\delta_i - \delta_j > 0$, dan zal de kans p_{ij} groter dan 0,5 zijn. Is daarentegen studie-onderdeel j zwaarder dan studie-onderdeel i , dus $\delta_i - \delta_j < 0$, dan zal p_{ij} kleiner dan 0,5 zijn. Willen wij tot een schatting van de schaalwaarden komen, dan zullen wij de kans p_{ij} als functie van $\delta_i - \delta_j$ nog nader moeten specificeren. Thurstone deed dit door te veronderstellen dat de kanslijn p_{ij} een cumulatieve normale verdeling is. Dit berust op de gedachte dat de subjecten de schaalwaarden δ_i op het conti-

Figuur 9.2. Fluctuaties van de waargenomen studietijd als normale verdelingen



nuüm niet exact kunnen vaststellen en de fluctuaties in de waargenomen studietijd van studie-onderdeel i normaal verdeeld is met gemiddelde δ_i . Voor twee studie-onderdelen i en j is dit in *figuur 9.2* weergegeven. Thurstone noemt dit proces, waarbij de feitelijke studietijd met een zekere onnauwkeurigheid wordt waargenomen, een 'discriminal process'. De nauwkeurigheid van waarneming wordt bepaald door de standaardafwijkingen σ_i en σ_j , die Thurstone aanduidt met 'discriminal dispersions'. In *figuur 9.2* wordt de studietijd van studie-onderdeel j veel nauwkeuriger waargenomen dan die van studie-onderdeel i. Men kan nu bewijzen dat het verschil in waargenomen studietijd van deze onderdelen ook normaal verdeeld is met gemiddelde $\delta_i - \delta_j$ en standaardafwijking $\sqrt{\sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}}$, waarbij σ_{ij} de covariantie tussen i en j is. Wil men tot schattingen van de parameters komen, dan moet men echter nog extra vereenvoudigingen aanbrengen. De meest vergaande vereenvoudiging is die waarbij men veronderstelt dat voor alle paren i en j de standaardafwijking $\sqrt{\sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}}$ gelijk is. Men mag dan evengoed zeggen dat zij alle gelijk aan 1 zijn. Thurstone duidt dit model met gelijke standaardafwijkingen aan met CASE V. Pas als dit model verworpen moet worden, gaat men een algemenere versie toetsen. Wij zullen ons echter tot CASE V beperken. De kans dat studie-onderdeel i zwaarder wordt waargenomen dan j hangt dan uitsluitend van het verschil $\delta_i - \delta_j$ af. De daarbij behorende cumulatieve kanslijn p_{ij} , die in *figuur 9.1* getekend is, is dan voor alle tweetallen exact gelijk. Hiermee is het model geheel gespecificeerd. De parameters van het model, de studietijden δ_i , worden geschat via een schatting van de p_{ij} . Hiervoor nemen wij de fractie van de ondervraagde studenten die studie-onderdeel i als zwaarder heeft waargenomen dan studie-onderdeel j. Deze geschatte fractie duiden wij aan met $\hat{p}_{ij}$. (De studenten worden in dit model dus uitsluitend als replicaties van elkaar gezien en gebruikt om tot een redelijke schatting van de p_{ij} te komen. Via een tabel voor de standaard-normale verdeling kunnen wij de $\hat{p}_{ij}$ omzetten in z-scores; deze z-scores kunnen wij nu onmiddellijk gelijk stellen aan het verschil in schaalwaarden δ_i en δ_j .)

$z_{ij} = \delta_i - \delta_j$. Dat het model restricties aan de data, de $\hat{p}_{ij}$ oplegt, kunnen wij eenvoudig zien, wanneer wij $\hat{p}_{ij}$, $\hat{p}_{jk}$ en $\hat{p}_{ik}$ met elkaar vergelijken. Stel dat $\hat{p}_{ij} = 0,66$. Dit komt overeen met een z-score van 0,40, m.a.w.:

$$\delta_i - \delta_j = 0,40.$$

Stel verder dat $\hat{p}_{jk} = 0,73$, dan is het verschil tussen δ_j en δ_k gelijk aan 0,60:

$$\delta_j - \delta_k = 0,60.$$

Uit het model volgt nu dat $\delta_i - \delta_k = 0,40 + 0,60 = 1,00$ en dat $\hat{p}_{ik}$ gelijk moet zijn aan 0,84, de kans die overeenkomt met een z-score van 1,00. Dit is nog eens in *figuur 9.3* weergegeven, die aan Dawes (1972, p. 9) is ontleend. Deze restrictie staat bekend onder de naam 'strong stochastic transitivity' (Coombs, Dawes en Tversky, 1970, p. 156).

Bij m stimuli, die paarsgewijze worden voorgelegd, verkrijgen wij $\frac{1}{2}m(m-1)$ gegevens $\hat{p}_{ij}$ en z_{ij} . We schatten nu de δ_i zodanig dat de uit de geschatte δ_i terugberekende z_{ij} zo weinig mogelijk van de z_{ij} uit onze gegevens verschillen (zgn. kleinste kwadratenoplossing). Uit deze z_{ij} kunnen wij weer kansen $\hat{p}_{ij}$ berekenen die uit ons model voortvloeien. Deze kunnen wij vergelijken met de $\hat{p}_{ij}$ die wij uit onze gegevens hebben berekend. Mosteller heeft een chi-kwadraat toets ontworpen om na te gaan of de verschillen tussen de $\hat{p}_{ij}$ en $\hat{p}_{ij}$ aan toeval mogen worden toegeschreven of niet.

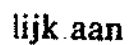
Dit is een van de belangrijkste toetsen, op basis waarvan wij beoordelen of het model met de daaruit berekende schattingen δ_i van de parameters acceptabel is of niet.

In het onderzoek van Rutges waren 9 studie-onderdelen uit de propedeuse en één studie-onderdeel, het computerpracticum, uit het tweede jaar betrokken. Het computerpracticum is eraan toegevoegd, omdat dat onderdeel in één week geconcentreerd wordt gegeven en zich uitstekend leent om de schaal te ijken, zoals wij hierna nog zullen zien. Om een passende oplossing te krijgen, moest één onderdeel uit de propedeuse worden weggelaten, namelijk 'Inleiding Wiskunde', dat uitsluitend voor studenten met een deficiënte vooropleiding is ingelast. Na weglating van dit onderdeel bleken de overige 9 studie-onderdelen te voldoen aan de veronderstellingen van het 'CASE V'-model (chi-kwadraat: 32,9, DF = 26, $p = 0,17$). De studie-onderdelen bleken dus op een intervalschaal geordend te kunnen worden (zie *tabel 9.1*, kolom 1). Het lichtste studie-onderdeel was het computer-

$$= 0,66.$$

net een
Dawes
'strong
)

edeuse
ar be-
onder-
nt om
oplos-
laten,
en de-
leken
n van
tudie-
orden
outer-

$$= 0,66.$$


net een
Dawes
'strong
)

twee extreme ijkpunten en de schaalwaarden vanuit het Thurstone-model de werkuren voor alle studie-onderdelen zijn berekend. Het aldus berekende aantal studie-uren voor het derde ijkpunt kon vergeleken worden met het door de studenten zelf geschatte aantal studie-uren. *Tabel 9.1* laat zien dat deze twee zeer redelijk met elkaar overeenkomen.

Tabel 9.1. Schaalwaarden en 'urenberekeningen' van de verschillende studie-onderdelen

studie-onderdelen	1 schaalwaarde ($\bar{z}$ -score)	2 gem. aantal uren van de ijkpunten	3 berekende uren op basis van kolom 1 en 2	4 controle berekeningen
Computerpracticum	0	26 uur	26 uur	
Inleiding Kansrek.	.853		87 uur	
Inleiding Statistiek	1.251		115 uur	112 uur
Sociologie II B	1.459		130 uur	
Sociologie II A	1.471		131 uur	
Thema-groep	1.489		132 uur	
Logica	1.667		145 uur	
Methodologie	2.435		200 uur	
Sociologie I	2.519	206 uur	206 uur	

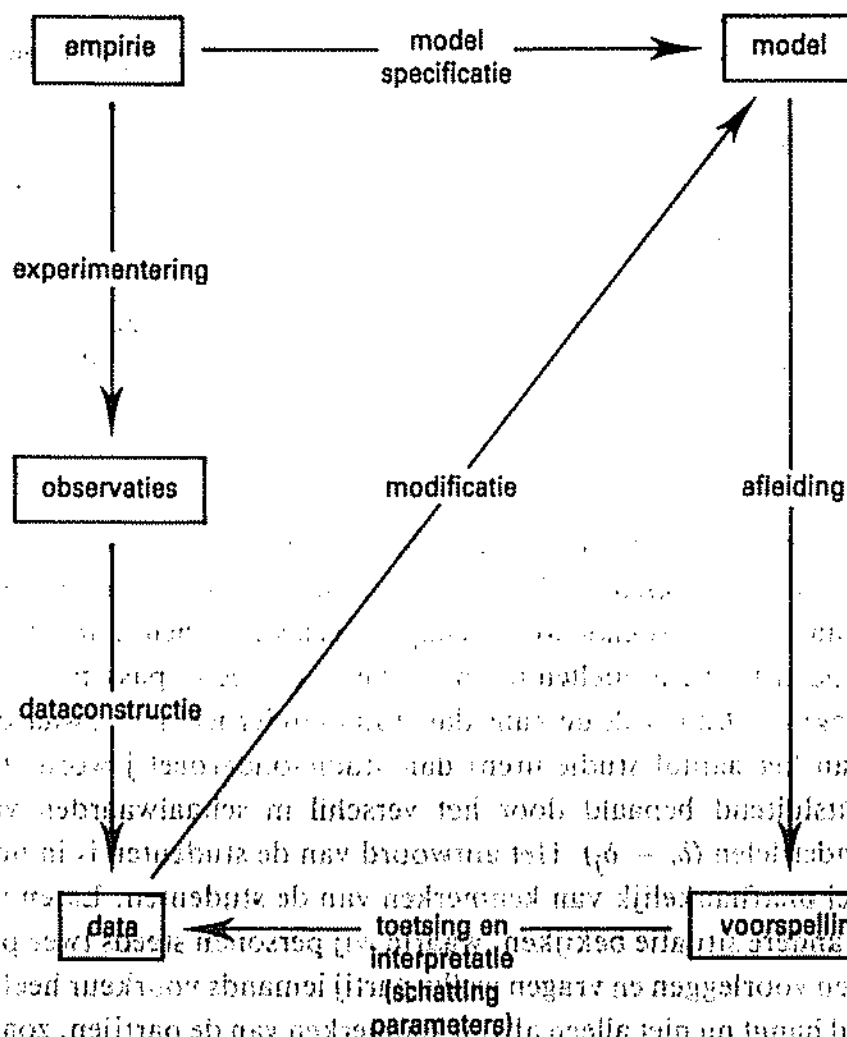
Bron: Rutges, 1978, p. 19.

9.3. BESLISSINGSMOMENTEN

In *figuur 9.4* is het meten en analyseren met modellen in sociologisch onderzoek nog eens schematisch weergegeven. Deze weergave is een verdere uitwerking van het schema van het wetenschappelijk onderzoeksproces, zoals dat door Coombs, Dawes en Tversky (1970, p. 3) is weergegeven. Het schema is zowel van toepassing op het meetproces met behulp van meetmodellen, waarvan wij in de vorige paragraaf een voorbeeld hebben uitgewerkt, als op het analyseren van de structuur, de samenhang tussen een aantal sociale kenmerken.

Bij de toelichting van *figuur 9.4* zullen wij ons vooral concentreren op meetmodellen, omdat aan de hand daarvan gemakkelijk kan worden gedemonstreerd hoe hetzelfde type gegevens op basis van alternatieve modellen geheel verschillend kan worden geïnterpreteerd. Wel zal zijdelings daarbij

Figuur 9.4. Schematische weergave van het meten en analyseren met modellen in sociologisch onderzoek



een aantal opmerkingen worden geplaatst over de toepassing van het schema op structurele modellen.

9.3.1. Modelspecificatie

Een meetmodel specificeert zowel de aard van het latente kenmerk als de wijze waarop het manifeste gedrag daaraan is gerelateerd. In het voorbeeld van paragraaf 9.2 wordt het latente kenmerk (studietijd) gespecificeerd als een ééndimensionale continue grootheid. Naast ééndimensionale modellen bestaan er echter ook meerdimensionale modellen. Meerdimensionale modellen zijn zeer geschikt, wanneer wij met samengestelde kenmerken te maken hebben. In meerdimensionale modellen worden objecten als punten in een twee- of meerdimensionale ruimte gezien. Voor elke dimensie krijgt

het object nu een schaalwaarde toegekend. Deze schaalwaarden komen overeen met de waarden van het object op de coördinatenassen van de meerdimensionale ruimte.

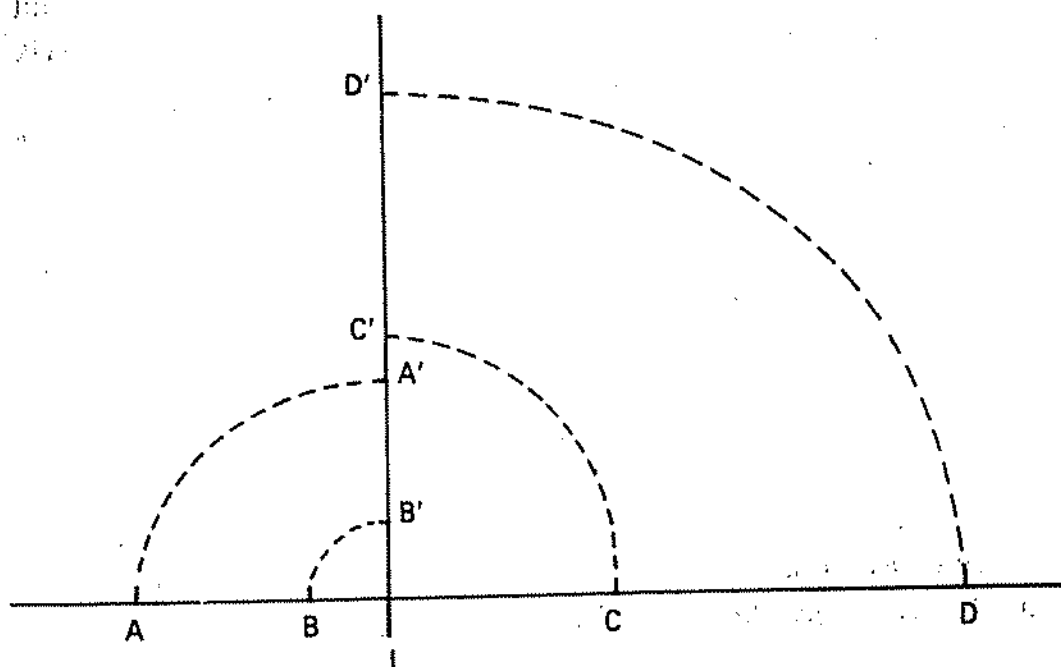
In het voorbeeld van paragraaf 9.2 is het manifeste gedrag via een kansfunctie aan het latente kenmerk gerelateerd (zie *figuur 9.1*). Zulke modellen heten stochastische modellen. In dergelijke modellen vormt de onnauwkeurigheid van het manifeste gedrag een integraal onderdeel van het model. Dit is bij deterministische modellen niet het geval; daarin komen uitsluitend kansen 0 of 1 voor. In een deterministisch model zou elke student minder werkuren moeten besteden aan een studie-onderdeel met een lagere schaalwaarde. Een deterministisch model is dan ook veel stringenter dan een stochastisch model. Een ander nadeel van deterministische modellen is dat uit het model geen statistische criteria zijn af te leiden om het model te toetsen. Bij deterministische modellen kunnen wij hoogstens het aantal verstoringen tellen en daarop bepaalde criteria formuleren om te beslissen of dit aantal zo klein is dat wij het model toch maar accepteren. Er bestaat daarom in de sociale wetenschappen een sterke behoefte, waar mogelijk, stochastische modellen te ontwikkelen en toe te passen.

In paragraaf 9.2 wordt de kans dat studie-onderdeel i als zwaarder (in de zin van het aantal studie-uren) dan studie-onderdeel j wordt beoordeeld, uitsluitend bepaald door het verschil in schaalwaarden van de studie-onderdelen ($\delta_i - \delta_j$). Het antwoord van de studenten is in het model geheel onafhankelijk van kenmerken van de studenten. Laten wij nu eens een andere situatie bekijken, waarin wij personen steeds twee politieke partijen voorleggen en vragen welke partij iemands voorkeur heeft. Het antwoord hangt nu niet alleen af van kenmerken van de partijen, zoals hun politieke stellingname, maar ook van de politieke stellingname van de persoon zelf. Immers voor een socialistische georiënteerde persoon zal de vraag naar de voorkeur van PvdA boven VVD geheel anders uitvallen dan voor een liberaal georiënteerde persoon. De personen mogen nu dan ook niet meer als replicaties van elkaar gezien worden. Wij hebben hier een geheel ander model nodig. Zo'n ander model zou bijv. het ontvouwingsmodel van Coombs (1963) kunnen zijn. Uitgaande van vier politieke partijen is dit in *figuur 9.5* gegeven. Wederom specificceert het model een ééndimensionaal kenmerk. Nu hebben echter niet alleen de stimuli (de politieke partijen) een ons onbekende schaalwaarde op het continuüm, maar ook de personen zelf. De schaalwaarde van een persoon komt overeen met zijn politieke ideaal.

Het model van Coombs postuleert nu dat de voorkeursrangorde van stimuli en dus ook de paarsgewijze voorkeur, bepaald wordt door de afstand

Figur
die
keur
hoo
dgi
waa
nuu
figu
van
voor
den
del g
wij j
scha
W
moe
serv
de t
stoc
ene
slui
moe
mul

Figuur 9.5. Ontvouwingsmodel van Coombs



die die stimuli hebben t.o.v. het ideaalpunt van een persoon. De voorkeursrangorde van de personen met ideaalpunt I in *figuur 9.5* zal dus van hoog naar laag zijn BACD, aangezien:

$$d_{BI} < d_{AI} < d_{CI} < d_{DI}$$

waarbij d_{BI} de afstand tussen stimulus B en ideaalpunt I is op het continuüm. Als wij het continuüm op het ideaalpunt vouwen (de verticale as in *figuur 9.5*), dan wordt de voorkeursrangorde bepaald door de rangorde van de stimuli op die verticale as. Het model laat zien hoe paarsgewijze voorkeuren of voorkeursrangorden bepaald worden vanuit de schaalwaarden van stimuli en personen op het continuüm. Bij toepassing van het model gaan wij natuurlijk, net als in paragraaf 9.2, de andere kant op, waarbij wij proberen vanuit de voorkeursrangorden een schatting te krijgen van de schaalwaarden van zowel stimuli als personen.

Waarin verschilt nu het model in *figuur 9.5* van dat in *figuur 9.1*? Wij moeten ons daarbij realiseren dat wij in beide gevallen dezelfde soort observaties hebben, nl. een paarsgewijze vergelijking van stimuli. Toch zijn de twee modellen fundamenteel verschillend. Het model in *figuur 9.1* is stochastisch. Het model in *figuur 9.5* is deterministisch. De keuze van de ene stimulus boven het andere wordt in het Thurstone-model verder uitsluitend bepaald door de kenmerken van de stimuli, terwijl in het Coombs-model de keuze bepaald wordt door kenmerken van zowel personen als stimuli. Het model in *figuur 9.1* kent dus maar één soort parameter dat in *fi-*

guur 9.5 twee. Tenslotte verschillen de twee modellen nog wezenlijk in een derde opzicht. De keuze tussen twee studie-onderdelen in *figuur 9.1* houdt een vergelijking door de student in tussen twee objecten: namelijk welk studie-onderdeel is zwaarder. De keuze naar voorkeur tussen twee politieke partijen A en B in *figuur 9.5* houdt echter een vergelijking in tussen twee paren objecten: vergeleken wordt immers de afstand tussen partij A en het ideaal I (d_{AI}) met de afstand tussen partij B en het ideaal I (d_{BI}).

Zoals bekend vormen de twee laatste verschillen tussen het Thurstone-model en het Coombs-model de basis voor een door Coombs opgestelde typologie van meetmodellen. Deze typologie kent vier fundamentele klassen van meetmodellen. Deze zijn in *tabel 9.2* gegeven met voor elke klasse een of meer namen van personen die dergelijke modellen ontwikkeld hebben. Voorbeelden van toepassingen van deze typen modellen kan men vinden in paragraaf 9.2 (het Thurstone-model), in hoofdstuk 10 (waar het Mokken en Rasch-model toegepast en met elkaar vergeleken worden) en in hoofdstuk 11 (gelijkenismodellen).

Tabel 9.2. Meetmodellen volgende de typologie van Coombs

relaties worden gelegd tussen paren objecten

aantal soorten parameters	modellen van paars- gewijze vergelijking (Thurstone; Bradley Terry Luce)	gelijkenismodellen (Torgerson; Kruskal)
	latente trekmodellen (Guttman; Mokken; Rasch)	(meerdimensionale) preferentie modellen (Coombs)

De latente trekmodellen in *tabel 9.2* vormen een aantrekkelijk alternatief voor het factoranalytische model, waarin uitgegaan wordt van een lineair verband tussen manifest gedrag en een of meer latente kenmerken, zoals hoofdstuk 10 laat zien voor diverse aspecten van schoolbeleving (zie ook Molenaar, 1982). Desalniettemin blijft het lineaire model van groot belang, als meetmodel maar vooral ook als structureel model, en heeft zelfs aan kracht gewonnen door de op maximalisering van de aannemelijkheidsfunctie gebaseerde schattings- en toetsingsprocedures, die door Jöreskog c.s. onder de namen confirmatorische factoranalyse (COFAMM) en lineair structural relations (LISREL) zijn ontwikkeld. In hoofdstuk 13 en 14 zijn hiervan interessante voorbeelden te zien. Van de Pol laat hierbij zien hoe

ook h
gebru
in bet
analy
proce
van h
tesam
vereis
len).
ding
raadz
REL
of sch
een a
ren. E
15.
9.3.2.
Een
paald
nenv
ties (2
van d
afbee
neel s
1971)
1. de
for
pr
2. de
ge
3. de
di
te
ke
Zoals
lijkin
verei
nier

ook het structurele model van LISREL bij herhaalde metingen kan worden gebruikt om betrouwbaarheid van metingen te schatten, zodat verschillen in betrouwbaarheid tussen predictoren bij een daarop volgende regressie-analyse kunnen worden geëlimineerd. Ormel illustreert hoe de LISREL-procedures kunnen worden toegepast voor toetsing en parameterschatting van het structurele model alleen, maar ook voor meet- en structureel model tesamen. Een belangrijk nadeel van deze modellen vormen de vrije strenge vereisten die zij stellen aan de data (continue en normaal verdeelde variabelen). De gevoeligheid van de schattings- en toetsingsprocedure voor schending van deze assumpties is onderzocht door Boomsma (1983). Het lijkt raadzaam deze studie in ieder geval te raadplegen bij toepassing van LISREL op kleine steekproeven en in situaties waarbij de variabelen discreet of scheef verdeeld zijn. In deze laatste gevallen vormt het loglineaire model een aantrekkelijk alternatief om de structuur tussen kenmerken te analyseren. Een toepassing hiervan op sociale indicatoren vindt men in hoofdstuk 15.

9.3.2. *Experimentering en dataconstructie*

Een meetmodel kan slechts ingeschakeld en getoetst worden indien wij bepaalde observaties hebben in een zodanige vorm dat zij in het model kunnen worden ingebracht. Een eerste stap daartoe is de keuze van de *observaties* (zie *figuur 9.4* en Coombs, 1963, 4) door experimentering. Uitgaande van de moderne meettheorie, waarin het meten wordt beschouwd als een afbeelding van een empirisch relationeel systeem in een numeriek relationeel systeem (zie bijv. Pfanzagl, 1968; Krantz, Luce, Suppes en Tversky, 1971), moeten daarbij de volgende drie keuzes gemaakt worden:

1. de keuze van de *objecten* als drager van de kenmerken: welke populatie onderzoeken wij? Onderzoeken wij de gehele populatie of een steekproef daaruit?
 2. de keuze van de *indicatoren* (de manifeste kenmerken) die geacht mogen worden betrekking te hebben op het latente kenmerk;
 3. de keuze van de vast te leggen *empirische relaties* tussen objecten en indicatoren, indicatoren onderling en, al dan niet daaruit afgeleid, objecten onderling. Dit wordt deels bepaald door het meetmodel dat wij denken in te schakelen.
- Zoals de naam reeds aangeeft, vereisen modellen van paarsgewijze vergelijking dat objecten paarsgewijze worden voorgelegd. Preferentiemodellen vereisen veelal voorkeursrangorden. Deze moeten op de een of andere manier uit onze observaties zijn af te leiden, hetzij doordat wij voorkeurs-

rangorden hebben gevraagd, hetzij vanuit paarsgewijze voorkeuren, hetzij vanuit voorgelegde triaden. De modellen van Guttman, Mokken en Rasch vereisen een aantal items (bijv. uitspraken) waarop de antwoorden te herleiden zijn tot een positief en een negatief alternatief. De gelijkenismodellen vereisen tussen elk tweetal objecten een maat (eventueel ordening) van 'gelijkenis'. Ook deze moet op de een of andere manier uit onze observaties zijn af te leiden. Wij zien dus dat onze observaties gedeeltelijk gekozen moeten worden vanuit het model, c.q. de modellen die wij toepasbaar achten. Uitdrukkelijk staat hier gedeeltelijk bij, omdat soms observaties geschikt gemaakt kunnen worden voor diverse soorten modellen. Vandaar dat de transformatie van onze observaties tot *data* via dataconstructie als een afzonderlijke stap in *figuur 9.4* is opgenomen. Data zijn analytisch geïnterpreteerde observaties, d.w.z. observaties in een zodanige vorm getransformeerd dat zij geschikt zijn om de parameters van het model te schatten en de restricties, die uit het model voortvloeien te toetsen. Voor multivariate analysemodellen bestaan de data uit de correlaties of covarianties tussen de variabelen. In paragraaf 9.2 hebben wij gezien dat voor het Thurstone-model de data bestaan uit de β_{ij} die direct uit de observaties van paarsgewijze vergelijkingen voortvloeien. De data voor het Coombs-model van *figuur 9.5* bestaan echter uit de voorkeursrangorden die direct kunnen worden gevraagd, maar ook bijvoorbeeld uit observaties van paarsgewijze vergelijkingen kunnen worden afgeleid. Het zal dan ook duidelijk zijn dat een scherp onderscheid gemaakt moet worden tussen paarsgewijze vergelijkingen als observaties en modellen van paarsgewijze vergelijking.

9.3.3. Afleiding, toetsing en interpretatie

Wil een model meer dan heuristische waarde hebben, dan moeten uit het model kunnen worden afgeleid:

1. *Restricties* waaraan de data moeten voldoen (zie *figuur 9.4*). Op basis van deze restricties kunnen *maten en criteria voor passing (goodness of fit)* worden gedefinieerd en gehanteerd. Goede passingsmaten en -criteria zijn van belang, omdat zij bepalen wanneer een model aanvaard of verworpen wordt. Sommige meetmodellen kennen dergelijke maten niet alleen voor de schaal als geheel, maar ook voor afzonderlijke items of stimuli om te beoordelen welke items wel en welke niet in een schaal passen. Dit is van belang om te localiseren waar het meetmodel niet klopt, en het meetmodel als hulpmiddel van itemselectie te gebruiken, d.w.z. als hulpmiddel om uit-

einde
het m
van
aang
dente
2. Sc
spec
struc
tisch
meet
vaard
verde
vallen
daarb
van h
gege
ven b
bestaa

Bij toe
van he
ters va
in het
gestra
het mo

9.3.4.

Indien
ons m
9.4)? B
lyse va
spoor l
opheff
resulta
meetm
model
model
vervang
niet om

eindelijk een goed meetinstrument te creëren. Zo bleek in paragraaf 9.2 dat het model niet voor alle studie-onderdelen opging, maar wel na eliminatie van 'Inleiding wiskunde'. Aangezien ook op andere gronden kon worden aangevoerd dat dit een afwijkend studie-onderdeel was (alleen voor studenten met deficiënte vooropleiding), was deze eliminatie acceptabel.

2. *Schattingen van de parameters* van het model. Bij structurele modellen specificeren de parameterschattingen de veelal als causaal geïnterpreteerde structuur tussen de kenmerken. Deze structuur zal in een ruimer theoretisch kader moeten worden geplaatst, zo die niet daaruit is afgeleid. Bij meetmodellen zijn deze schattingen de grootheden, waarmee wij, na aanvaarding van het model op basis van de onder 1 geformuleerde criteria, verder gaan werken. *Tabel 9.1* is daar een voorbeeld van. In een aantal gevallen kan het echter zinvol zijn wel het model te toetsen, maar niet de daarbij afgeleide schattingen van de parameters te gebruiken voor de rest van het onderzoek. Met name bij factoranalyse wordt vaak de voorkeur gegeven aan ongewogen optelling van de scores op de betreffende items boven berekening van factorscores, omdat er toch een zeer hoge correlatie bestaat tussen gewogen en ongewogen somscores.

Bij toepassingen van modellen kan het hoofddoel liggen op de toetsing van het model via de criteria van passing of op de schatting van de parameters van het model. In het ene geval gebruikt men het model als criterium, in het tweede als instrument. Het zal echter duidelijk zijn dat men niet ongestraft het model als instrument kan gebruiken wanneer de passing van het model zeer slecht is.

9.3.4. *Modificatie*

Indien het model wordt verworpen, welke modificaties brengen wij dan in ons model aan, m.a.w. welke modelassumpties wijzigen wij (zie *figuur 9.4*)? Bij structurele modellen geschiedt de aanpassing veelal door een analyse van de residuen, waardoor men fouten in de modelspecificatie op het spoor kan komen. Men kan dan bepaalde restricties in het model alsnog opheffen indien dit niet ten koste gaat van de interpreteerbaarheid van de resultaten. De hoofdstukken 13, 14 en 15 geven hier voorbeelden van. Bij meetmodellen staan verschillende wegen open om bij verwerping van het model modificaties aan te brengen. Ten eerste kunnen wij nagaan of het model passend is te maken door één of meer stimuli te laten vallen of te vervangen. Op deze wijze is in paragraaf 9.2 gehandeld. Gaat het hierbij niet om een incidentele stimulus of incidenteel item, maar vermoedelijk om

een gehele groep van stimuli, dan kunnen wij proberen het model voor subgroepen van stimuli apart toe te passen. Deze handelwijze wordt vaak aangeduid met 'multiple scaling'. Zo blijkt het model in *figuur 9.5* niet op te gaan voor de Nederlandse politieke partijen. Het is denkbaar dat het enerzijds voor de confessionele partijen en anderzijds voor de niet-confessionele partijen apart wel opgaat. Voorbeelden van deze benaderingswijze vindt men in hoofdstuk 11. Een tweede modificatie kan bestaan uit het weglaten van subjecten. Wij zouden bijv. studenten die intransitief hebben gekozen, kunnen verwijderen en het model uitsluitend op de overige studenten toepassen. Gaat het niet om incidentele gevallen, maar vermoedelijk om een gehele groep van subjecten, dan kunnen wij proberen het model voor elke subgroep afzonderlijk toe te passen. Als bijvoorbeeld studenten met een wiskunde I-vooropleiding een fundamenteel andere rangorde van studie-onderdelen zouden hebben gehad dan de overige studenten, zouden wij wellicht twee schaalwaarden hebben moeten accepteren: een voor de eerste subgroep van studenten en een voor de tweede. Een derde modificatie kan bestaan uit het kiezen van een minder stringent model. Past een deterministisch model niet, dan kunnen wij wellicht een vergelijkbaar stochastisch model toepassen. Helaas zijn voor het model in *figuur 9.5* nog geen goede stochastische alternatieven gevonden. Wel zijn verschillende stochastische modellen geformuleerd, maar aan de parameterschatting en goede passingscriteria wordt nog (o.a. in enkele Nederlandse projecten) gewerkt. Een andere mogelijke versoepeling van een meetmodel bestaat in het toepassing van een meerdimensionaal model als een ééndimensionaal model niet blijkt te passen. Voor het ééndimensionale model in *figuur 9.5* is een vergelijkbaar meerdimensionaal model inderdaad aanwezig. Zo blijken de Nederlandse politieke partijen en de ideaalpunten van kamerleden wel in een tweedimensionaal model te passen, waarbij de rangorde van afstanden van een ideaalpunt naar de politieke partijen overeenkomt met de voorkeursrangorde van het betreffende kamerlid (Daalder en Rusk, 1972, p. 175 e.v.; zie ook Daalder en Van der Geer, 1977).

9.4. SLOT

In de vorige paragraaf is het gehele proces van meten en analyseren met modellen voor sociologisch onderzoek gesystematiseerd en zijn de verschillende beslissingsmomenten daarin expliciet aangegeven. Mede aan de hand van een concreet voorbeeld (zie paragraaf 9.2) is duidelijk gemaakt dat het ad hoc-karakter van metingen en analyses in de sociale we-

tensch:
worde
verkla
doelste
rieën
meeste
model
singsp
schijn
afzonn
sing, v
versch
mulati

Literat

Bezeml

tastr

Booms

mali

Coomt

Coomt

eme

Daalde

Gen

** Daa

S.C.

tiers

Dayes

** Gad

caus

1982

Jöresk

meti

Jöresk

estir

ceea

Jun

Krantz

men

pres

Krusk

met

Meerli

obs

tenschappen door een dergelijke modelmatige aanpak aanmerkelijk kan worden teruggedrongen (zie ook Dawes, 1972 en Molenaar, 1982). Met Verklarende modellen in deel I van deze bundel wordt een soortgelijke doelstelling beoogd: primair het ad hoc-karakter bij de opstelling van theorieën en structurele modellen te elimineren. In beide situaties wordt meestal een groot aantal theoretische beweringen samengevoegd en tot één model uitgewerkt, dat vervolgens integraal kan worden getoetst. De toetsingsprocedure is daarbij vooral gericht op het totaalbeeld van een verschijnsel – het model als geheel –, meer dan op vrij los van elkaar staande afzonderlijke hypothesen (Meerling, 1980, p. 33-34). Ontwikkeling, toetsing, verbetering en eventueel verwerping van modellen voor sociologische verschijnselen wordt daarmee een belangrijke motor in het proces van cumulatieve kennisverwerving.

Literatuur

- Bezembinder, Th.G.G., *Van rangorde naar continuum. Een verhandeling over datastructuren in de psychologie*. Deventer: Van Loghum Slaterus, 1970.
- Boomsma, A., *The robustness of LISREL against small sample size and nonnormality*. Proefschrift Groningen, te verschijnen in 1983.
- Coombs, C.H., *A theory of data*. New York: Wiley, 1964.
- Coombs, C.H., R.M. Dawes, and A. Tversky, *Mathematical psychology. An elementary introduction*. Englewood Cliffs: Prentice Hall, 1970.
- Daalder, H., en J.P. van der Geer, Partijafstanden in de Tweede Kamer der Staten-Generaal. *Acta politica*, 12, 1977, p. 289-345.
- ** Daalder, H., and J.G. Rusk, Perceptions of party in the Dutch parliament. In: S.C. Patterson and J.C. Wahlke (eds.), *Comparative legislative behavior. Frontiers of research*. New York: Wiley, 1972, p. 143-198.
- Dawes, R.M., *Fundamentals of attitude measurement*. New York: Wiley, 1972.
- ** Gadourek, I., *Social change as redefinition of roles. A study of structural and causal relationships in the Netherlands of the 'seventies'*. Assen: Van Gorcum, 1982.
- Jöreskog, K.G., A general method for analysis of covariance structures. *Biometrika*, 57, 1970, p. 239-251.
- Jöreskog, K.G., Structural equation models in the social sciences. Specification, estimation and testing. In: P.R. Krishnaiah (ed.), *Applications of statistics. Proceedings of the symposium held at Wright State University, Dayton, Ohio, 14-18 June 1976*. Amsterdam: North-Holland publishing company, 1977, p. 265-287.
- Krantz, D.H., R.D. Luce, P. Suppes, and A. Tversky, *Foundations of measurement. Vol. I: Additive and polynomial representations*. New York: Academic press, 1971.
- Kruskal, J.B., Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika*, 29, 1964, p. 1-27.
- Meerling, *Methoden en technieken van psychologisch onderzoek. Deel I: Model, observatie en beslissing*. Meppel: Boom, 1980.

- * Meerling, *Methoden en technieken van psychologisch onderzoek. Deel 2: Data-analyse en psychometrie*. Meppel: Boom, 1981.
- ** Mokken, R.J., *A theory and procedure of scale analysis. With applications in political research*. The Hague: Mouton, 1971.
- * Molenaar, I.W., Mensen die het beter meten. Een inleiding tot enige latente trek modellen. *Kwantitatieve methoden*, 3, 1982, nr. 5, p. 13-29.
- Pfanzagl, J., *Theory of measurement*. Würzburg: Physica-Verlag, 1968.
- Rutges, A., *Onderzoek naar de studiebelasting van studieonderdelen in het propedeusejaar 1976/77 bij de sociologieopleiding te Groningen*. Groningen: Sociologisch Instituut van de Rijksuniversiteit, 1978. (Bulletin van de Vakgroep Methoden en Technieken, 28)
- * Stokman, F., and W. van Schuur, Basic scaling. *Quality and quantity*, 14, 1980, p. 5-30.
- Thurstone, L.L., A law of comparative judgment. *Psychological review*, 34, 1927, p. 273-286.
- Torgerson, W.S., *Theory and methods of scaling*. New York: Wiley, 1958.

Bezoekadres: T.H. 1, Postbus 1000, 3720 BA Utrecht
 Telefoon: (043) 2511111
 Telefax: (043) 2511112
 E-mail: i.w.molenaar@uu.nl
 Website: www.uu.nl/~i.w.molenaar

- * inleidingen
- ** interessante toepassingen

10.
 I.W.
 Red:
 In h
 die v
 ving
 derz
 gen.
 zelf
 de k
 zaak
 tie v
 derz
 de a
 kan
 pen
 gaat
 en g
 toep
 mod
 scho
 item
 eiser
 antw
 Mok
 mee
 pers
 voor
 in sp
 pers
 toek
 de o
 den
 Juis
 ook
 dach
 rele
 wor