

KG-publicatie nr. 14
Schaalconstructie

Arie van Peet

KG-14

A. van Peet
Schaalconstructie

Correspondentie over deze KG-publicatie kunt u sturen naar:
E-mail: a.a.j.vanpeet@uva.nl

KG-publicaties bevatten interne notities, verslagen en prepublicaties bestemd voor spreiding in kleine kring.
De verantwoordelijkheid voor de inhoud van een KG-publicatie berust bij de auteur(s). Uit KG-publicaties mag
alleen geciteerd worden met toestemming van de auteurs.

Voorwoord

Ook in toegepast onderzoek worden veel attitudeschalen afgenomen en veel van deze schalen zijn nieuw. Het nu volgende wil een inleiding geven in drie schaaltypen, te weten: Likertschaal, Thurstoneschaal en de Guttmanschaal. Van deze drie schaaltypen is de Likertschaal verreweg de meest toegepaste.

Deze inleiding is gebaseerd op collegeaantekeningen voor tweede en derde jaars studenten in de opvoedkunde in het kader van een onderwijsonderdeel over testleer. De eerste vier paragrafen kunnen gelezen worden zonder enige kennis van de statistiek. De daaropvolgende paragrafen veronderstellen enige kennis over het gemiddelde, de standaardafwijking, de correlatiecoëfficiënt, Cronbachs alfa en principale componentenanalyse.

Commentaar en suggesties ter verbetering zijn welkom.

Arie van Peet

-
- nr.1 Everaert, H.A. en J.C. van der Wolf (2005).
 Behaviorally Challenging Students and Teacher Stress.
-
- nr.2 Wolf, J.C. van der & J.M.F. Touw (2005).
 Onderzoek naar zorg in het curriculum van de Theo Thijssen Academie.
-
- nr.3 Doorn, E.C. van (2005).
 De Gedragingenlijst voor Leraren (Tweede onderzoeksrapport).
-
- nr.4 Wolf, J.C. van der (2005).
 Probleemouders en de school: een onderontwikkeld terrein.
-
- nr.5 Enthoven, M. (2005).
 The contribution of the school environment to youths' resilience: A Dutch middle-adolescent perspective.
-
- nr.6 Enthoven, M.; A.C. Bouwer; J.C. Van der Wolf & A. van Peet (2005).
 Recognizing Resilience: Development and Validation of an Instrument to Recognize Resilience in Dutch Middle-Adolescents.
-
- nr.7 Velderman, H & H.A. Everaert (2005).
 Time-out or switch? (Paper presented at the ECER conference On 9 September 2005, University College Dublin).
-
- nr.8 Touw, J.M.F., J.T.E. van Beukering & H.A. Everaert (2005).
 Teachers' Personal Constructs on Problem Behaviour (Paper presented at the annual meeting of the European Educational Research Association (EERA), Dublin, Ireland, September 7-10).
-

nr.9 Doorn, E.C. van (2005).
Levend leren: daar ga ik voor!

nr.10 J.T.E. van Beukering, J.M.F. Touw & H. Everaert (2005).
Teachers' personal constructs on problem behaviour: towards professional development
& Kos, P. (2005). Personal constructs on (problem) pupils: a teacher's view.
*2 Papers presented at the International Practitioner Research Conference & Collaborative
Action Research Network Conference (PRAR 2005), Utrecht, The Netherlands,
November 4-6, 2005.*

nr. 11 Everaert, H. & A. van Peet (2006). Kwalitatief en kwantitatief onderzoek.

Nr. 12 Everaert, H.A. & J.C. van der Wolf (2006).
A Comparison of Stress and Burnout between Dutch General and Special Education
Teachers. Paper presented at the annual meeting of the American Educational Research
Association (AERA), San Francisco, USA, April 7-11, 2006.

Nr. 13 Everaert, H.A. & J.C. van der Wolf (2006). Gender Perceptions of Challenging Student
Behavior and Teacher Stress.

Nr. 14 Peet, A.A.J. van (2006). Schaalconstructie.

Inhoudsopgave

1	<i>Verschil tussen index en schaal</i>	6
2	<i>Het begrip 'attitude'</i>	8
3	<i>Wat is een 'schaal'?</i>	10
4	<i>Likertschaal</i>	16
5	<i>Coëfficiënt α en item-totaalcorrelaties</i>	21
6	<i>Voorbeeld schaalanalyse</i>	29
7	<i>Thurstoneschaal (methode van gelijkverschijnende intervallen)</i>	34
8	<i>Guttmanschaal (methode voor cumulatieve schalen)</i>	41
9	<i>Slotwoord schalen</i>	51
	<i>Literatuur</i>	54

1 Verschil tussen index en schaal

Aan meningen, opinies en attitudes van mensen wordt in onze samenleving veel belang gehecht en vandaar ook dat in de sociale wetenschappen opiniepeilingen en attitudemetingen een belangrijke rol spelen.

Er zijn vele methoden om attitudes te meten. Attitudes worden gemeten door middel van attitudeschalen. Voorbeelden van pedagogische attitudeschalen zijn attitudes ten aanzien van:

- belonen en straffen
- ouderlijk toezicht op de activiteiten van kinderen
- het hanteren van 'angst' als een middel om gedrag van kinderen te sturen
- kinderen die liegen
- traditionele gezinsideologie
- mate waarin kinderen door hun ouders acceptierend en positief worden tegemoet getreden
- mate van vrijheid die ouders hun kinderen toestaan,

Alvorens we nader ingaan op het begrip 'schaal' is het zinvol om eerst het verschil uiteen te zetten tussen de begrippen 'index' en 'schaal'. Deze begrippen worden nogal eens door elkaar gebruikt.

Een schaal is een maat waarmee een onderzoeker de intensiteit, richting, niveau of potentie van één begrip tracht te meten. Een schaal rangschikt responsen (of algemener gesteld: observaties) op één onderliggend continuüm. Een schaal kan berusten op één enkele indicator (vragen, items, enz.) of op meerdere indicatoren. Hierover meer in de volgende paragrafen.

Een index is een maat waarmee de onderzoeker verschillende uiteenlopende indicatoren combineert in één enkele score. Deze samengestelde score is vaak een eenvoudige som van multiple indicatoren. Van een construct worden verschillende componenten gemeten en gecombineerd tot één enkele maat.

Voorbeeld

Neem een tentamen bestaande uit 25 vragen over verschillende onderdelen van een cursus. Het aantal goed beantwoorde vragen is een soort index. Het is immers een samengestelde maat waarin elke vraag een klein stukje kennis en inzicht meet. Zou de docent echter de pretentie hebben om één onderliggend begrip te meten dan zou men van een schaal moeten spreken.

Voorbeeld

Een ander voorbeeld van een index is de index voor de kwaliteit van een faculteit. Deze index zou kunnen bestaan uit de volgende items:

- 1) aantal studenten per medewerker
- 2) percentage gepromoveerde medewerkers
- 3) aantal boeken in de bibliotheek per student

- 4) percentage eerste jaars dat afstudeert
- 5) percentage studenten dat na het behalen van de propedeuse afstudeert
- 6) aantal publicaties per medewerker

Men zou een aantal faculteiten op deze punten kunnen scoren om vervolgens per faculteit de scores op elk item bij elkaar op te tellen om also een indexscore te verkrijgen waarmee men faculteiten zou kunnen vergelijken.

Indexen kunnen met elkaar gecombineerd worden. We zouden bovenstaande index voor de kwaliteit van een faculteit kunnen verbeteren door er bijvoorbeeld een (sub-)index over de kwaliteit van het onderwijs aan toe te voegen, bestaande uit:

- a) gemiddelde groeps grootte,
- b) percentage lestijd gewijd aan discussie,
- c) aantal groepen per docent,
- d) beschikbaarheid van docenten buiten de lessen om,
- e) hoeveelheid literatuur die een student moet lezen,
- f) mate waarin de gegeven opdrachten het leren bevorderen,
- g) mate waarin de faculteit elke student kent,
- h) studentenoordelen over het lesgeven,

Een belangrijke vraag bij de constructie van een index is altijd of de samenstellende items al of niet gewogen moeten worden. Men zou bijvoorbeeld 'het aantal studenten per medewerker' belangrijker kunnen vinden dan 'het aantal boeken per student in de bibliotheek'; in dit geval moet men aan 'het aantal studenten per medewerker' een groter gewicht toekennen¹.

In §2 omschrijven we het begrip 'attitude' om daarna in §3 het begrip 'schaal' te omschrijven, alsmede het verschil tussen enkelvoudig meten en meervoudig meten. In de daaropvolgende paragrafen 4 t/m 8 behandelen we drie manieren om attitudes te meten: de Likertschaal, de Thurstoneschaal en de Guttman-schaal. Van deze drie is de Likertschaal het meest gebruikte schaaltype. In §9 tenslotte zullen we de drie schaaltechnieken kort met elkaar vergelijken.

¹ *De items van een index moeten vaak eerst gestandaardiseerd worden voordat zij gecombineerd kunnen worden. Een index behoeft dus niet persé een hoge coëfficiënt alfa te hebben, terwijl van een schaal wél een hoge alfa verwacht wordt.*

2 Het begrip 'attitude'

Zoals veel psychologische variabelen is een attitude een hypothetische of latente variabele; het is dus een niet direct waarneembare variabele. Het begrip 'attitude'² verwijst niet naar één handeling of antwoord van een persoon, maar het is een abstractie van een groot aantal handelingen en antwoorden. Wanneer we bijvoorbeeld stellen dat persoon A een minder gunstige attitude heeft ten aanzien van socialisme dan persoon B, dan bedoelen we dat de vele uitspraken, opmerkingen en handelingen van persoon A over het socialisme minder gunstig zijn dan de overeenkomstige uitspraken, opmerkingen en handelingen van persoon B. We mogen een samenbindend begrip als 'attitude' gebruiken als de vele, samenhangende antwoorden consistent zijn. Wanneer de mensen die bijvoorbeeld tegen een grotere vrije marktwerking zijn, ook degenen zijn die tegen privatisering van spoorwegen en openbare nutsbedrijven zijn en tevens degenen zijn die een grotere invloed van de overheid propageren, dan is het redelijk om te spreken van een houding die gunstig is ten aanzien van het socialisme.

In algemene termen gesteld: een latente variabele wordt gebruikt om de consistentie of covariatie te beschrijven tussen een aantal verschillende responsen op stimuli. De responsen covariëren (hangen samen) omdat ze gemedieerd worden door dezelfde hypothetische, latente variabele. De latente variabele verenigt een verzameling, namelijk de geobserveerde responsen. Andere termen voor een latente variabele zijn: trek (trait), interveniërende variabele, hypothetische variabele, factor, genotype. Andere namen voor 'geobserveerde variabele' zijn: manifeste variabele, phenotype en indicator.

Er zijn vele omschrijvingen van het begrip 'attitude' mogelijk. Eén daarvan is die van Campbell uit 1950:

"An individual's social attitude is an enduring syndrome of response consistency with regard to a set of social objects." (Campbell 1950, p.26).

In deze definitie wordt de nadruk gelegd op een kenmerk dat aan alle attitude-meting ten grondslag ligt, namelijk response-covariatie. Een iets andere omschrijving van het begrip 'attitude' is: "de neiging om consistent in gunstige of ongunstige zin te reageren op een object". Een attitude is altijd gericht op een object (een persoon, een situatie, instantie of gebeurtenis). Zo kan men spreken van attitudes ten aanzien van moderne meubelen (object), voetbalsupporters (personen), de overheid (instantie) en milieurampen (gebeurtenis).

De verzameling gedragingen, beweringen, etc. die tezamen de attitude vormen, heet het attitude-universum (de populatie). De elementen van dit attitude-universum zijn dus manifeste variabelen: de responsen op specifieke situaties (bijv. reacties op beweringen in termen van eens-oneens).

² Een toegankelijk boek dat geheel gewijd is aan de theorie en de meting van attitudes en opinies, is: Pligt, J.v.d. & Vries, N.K. de (1995). *Opinies en Attitudes (Meting, modellen en theorie)*. Amsterdam, Meppel: Boom.

Een steekproef van deze elementen wordt gebruikt om de attitude te meten. Begripsmatig gezien, is attitudemeting het trekken van een steekproef uit een universum. Dit betekent dat de steekproef representatief moet zijn. In de praktijk hebben we zelden de beschikking over het universum, maar hebben we alleen de beschikking over een steekproef. We trachten dan het universum te karakteriseren op grond van de resultaten van de steekproef waarbij het belangrijk is om niet te overgeneraliseren.

Op grond van verbale responsen op bijv. opinie vragen kunnen we niet zonder meer generaliseren naar niet-verbale situaties. Het kan zijn dat de antwoorden op de verbale vragen samen hangen met responsen in niet-verbale situaties maar dat moet dan wel empirisch vastgesteld worden. Anders gezegd: een persoon met een bepaalde attitude hoeft zich niet per se overeenkomstig die attitude te gedragen.

Zoals 'attitude' hiervoor is gedefinieerd, is het een latente variabele aangezien de betekenis wordt afgeleid uit de samenhang tussen responsen in een of ander attitude-universum. Als we een preciezere definitie van 'attitude' zouden willen hebben, hebben we een mathematisch model nodig dat de responsen (de observeerbare variabelen) relateert aan de latente variabele. Het probleem van het meten van een attitude is dus het bedenken van een (schaal-)model dat response-gegevens relateert aan de attitude-variabele. De Likertschaal, de Thurstoneschaal en de Guttmanschaal zijn voorbeelden van zulke schaalmodellen. We zullen in de nu volgende paragraaf eerst het begrip 'schaal' behandelen alvorens op deze drie schaalmodellen in te gaan.

3 Wat is een 'schaal'?

Een schaal is een techniek om houdingen, meningen, standpunten van mensen te meten. We kunnen een schaal omschrijven als een verzameling stimuli met betrekking tot één object. Zo bevat een 'attitudeschaal over zelfevaluatie' een aantal uitspraken over zelfevaluatie waarop personen antwoord geven.

Elke uitspraak is voorzien van twee of meer antwoordmogelijkheden(-categorieën), bijvoorbeeld:

- a. ja vs nee
- b. eens vs oneens
- c. op mij van toepassing vs niet op mij van toepassing'
- d. juist of niet juist
- e.
 - 1) precies op mij van toepassing
 - 2) grotendeels op mij van toepassing
 - 3) min of meer van mij van toepassing
 - 4) in beperkte mate op mij van toepassing
 - 5) helemaal niet op mij van toepassing
- f.
 - 1) geheel oneens
 - 2) grotendeels oneens
 - 3) enigszins oneens
 - 4) enigszins eens
 - 5) grotendeels eens
 - 6) geheel mee eens

Het aantal antwoordcategorieën ligt meestal tussen de vier en de acht. Er is geen vast voorgeschreven aantal, maar meer dan 10 antwoordcategorieën zal men niet vaak tegenkomen. De betrouwbaarheid stijgt niet meer wanneer men meer dan 10 categorieën heeft.

Een argument voor een oneven aantal antwoordcategorieën is dat de middelste categorie een neutrale, een 'middenpositie'-houding zou representeren. Maar ook een even aantal antwoordcategorieën komt voor (zie: Pligt 1995; Swanborn 1988).

Een 'schaal' werd hiervoor omschreven als een verzameling uitspraken. Waarom meerdere uitspraken en waarom niet slechts één uitspraak? Bijvoorbeeld één uitspraak over 'woonsatisfactie': Wat vindt U van de buurt waarin u woont? De bezwaren tegen het stellen van slechts één vraag zijn:

1. Heterogeniteit van het object

Het object is hier 'woonsatisfactie'. Welnu, het begrip 'woonsatisfactie' is zo heterogeen, beslaat zo veel verschillende aspecten, dat dit niet met één vraag te dekken valt.

Persoon A denkt bij de beantwoording aan het ontbreken van een parkeerplaats, persoon B denkt aan de ruzie met de buren, persoon C denkt aan het mooie uitzicht, persoon D denkt aan de supermarkt die vlak bij is en persoon-E denkt aan de verkeersdrukte die diezelfde supermarkt

aantrekt. Niet elke persoon zal alle aspecten rustig de revu laten passeren om vervolgens zijn antwoord te geven. Door het stellen van slechts één vraag kan men krijgen dat

- mensen met eenzelfde 'overall'-mening een verschillend antwoord geven
- mensen met een verschillende mening eenzelfde antwoord geven.

2. Toevalsfluctuaties

Allerlei niet er zake doende toevalsinvloeden kunnen een rol spelen: de persoon kan even afgeleid zijn, of hij kan de vraag verkeerd begrepen hebben of hij kan een opwelling van nijd hebben of juist even tolerant zijn. Veel van dit soort invloeden vallen tegen elkaar weg wanneer er meer vragen gesteld worden.

3. Systematische fouten

Wanneer er slechts één vraag gesteld wordt, oefenen systematische fouten een grotere invloed uit dan wanneer men meerdere vragen stelt. Voorbeelden van systematische fouten zijn:

- **Halo-effect**
De neiging om een overall positieve of negatieve beoordeling te geven. Docenten zijn bijvoorbeeld geneigd om van goede leerlingen meer fouten over het hoofd te zien dan van zeer slechte leerlingen.
- **Generositeitseffect**
De neiging om gunstige aspecten meer te benadrukken.
- **Vermijding van extreme categorieën**
De neiging om de uiteinden van een beoordelingsschaal niet te gebruiken.
- **Contrasteffect**
De neiging om anderen met jezelf te vergelijken.

Een object zal dus vaak te veel betekenissen hebben om het met één uitspraak te kunnen omschrijven. Dat geldt bijvoorbeeld ook voor het begrip 'Locus of Control'. Dit begrip kan men het best omschrijven door de twee polen van het begrip te omschrijven:

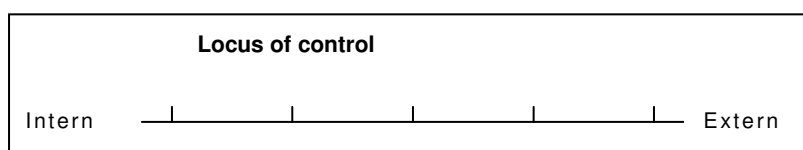
a) *Interne locus of control*

Respondenten met een sterke interne 'locus of control' menen dat vooral hun eigen gedrag (hun eigen kenmerken) de gebeurtenissen in hun leven bepalen.

b) *Externe locus of control*

Respondenten met een sterke externe 'locus of control' menen dat de gebeurtenissen in hun leven vooral bepaald worden door omstandigheden buiten henzelf, zoals toeval of de invloed van andere personen.

Stel dat de vraag aan een respondent zou zijn: "Waar plaatst U zichzelf op onderstaande lijn qua Locus of Control?"



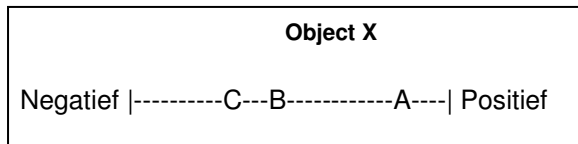
Ook voor dit begrip geldt dat verschillende respondenten bij het plaatsen van hun kruisje denken aan verschillende aspecten, bijvoorbeeld:

- de meeste narigheden in het leven overkomen je zonder dat je daar veel aan kunt doen.
- je stem uitbrengen bij verkiezingen is zeer belangrijk.

Wanneer er meer uitspraken over het begrip locus of control gepresenteerd worden, wordt de invloed van die min of meer toevallige associaties van de respondent bij het begrip minder en wordt de dekking van het begrip groter.

Het stellen van één enkele vraag heet enkelvoudig meten en het stellen van meerdere vragen heet meervoudig meten. Aan het stellen van meerdere vragen ligt het volgende basisidee ten grondslag.

Wanneer we het hebben over uiteenlopende meningen die respondenten ten aanzien van een object kunnen hebben, hebben we het in feite over een continuüm dat loopt van zeer negatief tot zeer positief. Op dit continuüm, dat we met een rechte lijn kunnen aangeven, bevinden zich de respondenten van wie we de positie zo precies mogelijk te weten willen komen.



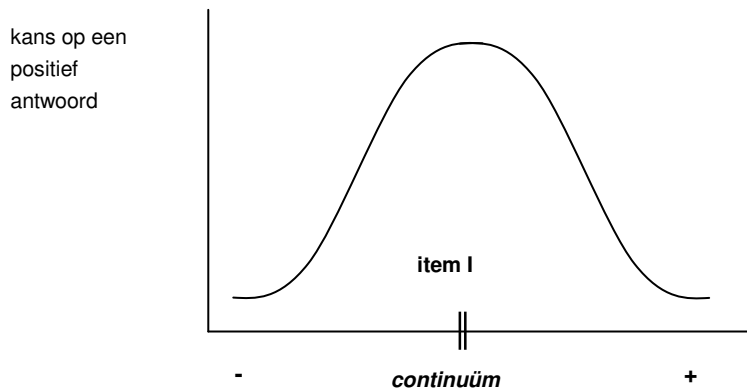
Respondent A heeft de meest gunstige attitude ten aanzien van object X, respondent B neemt een middenpositie in en respondent C heeft ten aanzien van object X een negatieve attitude.

We hebben het hiervoor een aantal keren gehad over uitspraken, beweringen, en dergelijke. De algemene term is stimulus of item. Ook items kunnen -evenals respondenten- onderscheiden worden naar hun positie op het continuüm. Er zijn positieve, neutrale en negatieve items. Een uitspraak als 'Op de tv zijn vaak boeiende programma's te zien' is duidelijk positiever over televisie dan een uitspraak als 'De programma's op de tv zijn meestal oppervlakkig'.

In deze behandeling van schalen gaan we er van uit dat zowel respondenten als items op één continuüm kunnen worden afgebeeld: ééndimensionale schaaltechnieken. Het is natuurlijk mogelijk en het komt ook vaak genoeg voor dat de attitude van een respondent meerdere dimensies bevat. Die meerdimensionale schaaltechnieken vallen buiten het kader van deze verhandeling.

Als zowel respondenten als items op eenzelfde continuüm kunnen worden geplaatst, rijst de vraag hoe deze zich tot elkaar verhouden. Wanneer reageert een persoon met een positief antwoord op een item? Een eenvoudig model wordt hieronder geschetst.

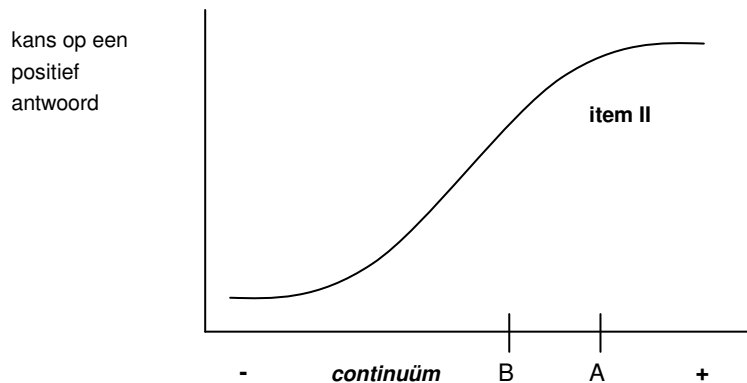
Figuur 1: Karakteristieke curve van een non-monotoon item



In bovenstaande figuur wordt aangegeven voor respondenten die zich op een bepaalde plaats van het continuüm bevinden, hoe groot de kans op een positief antwoord is. We gaan er gemakshalve even van uit dat men op item I slechts met nee of ja kan reageren. Respondenten die op het continuüm eenzelfde positie innemen als item I, hebben de grootste kans om 'ja' te zeggen. Naarmate de positie van de respondent verder van de positie van item I afligt, wordt de kans kleiner dat de respondent 'ja' zegt op het item. De lijn die voor item I en voor alle respondenten het verband aangeeft tussen de positie van een respondent op het continuüm en de kans voor die respondent om 'ja' te zeggen bij dat item, heet de itemkarakteristieke curve ('trace line'). Het hier geschetste verband tussen de positie van een respondent is in feite al een schaalmodel. We spreken hier van punt-items ofwel van non-monotone items: elk item kenmerkt zich door een punt op het continuüm ter weerszijde waarvan de kans op een positief antwoord afneemt.

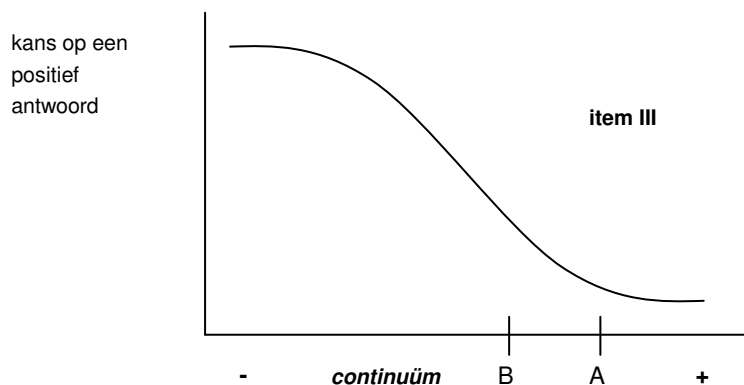
Naast non-monotone items zijn er ook monotone items. Een monotone item-karakteristiek wil zeggen dat kans op een positief antwoord voortdurend toeneemt, naarmate de positie van de respondent op het continuüm positiever is. Een voorbeeld van een monotone itemkarakteristieke curve wordt hierna gegeven.

Figuur 2: Karakteristieke curve van een monotoon item



Item II is een positief geformuleerd item: een respondent A die rechts van respondent B ligt heeft een grotere kans om 'ja' te zeggen dan respondent B. Bij een negatief geformuleerd item zou een itemkarakteristieke functie er als volgt uit kunnen zien.

Figuur 3: Karakteristieke curve van een monotoon item (negatief geformuleerd item)



Bij een negatief geformuleerd item heeft een respondent A die rechts van respondent B ligt een kleinere kans om 'ja' te zeggen dan respondent B.

Welke items hebben een monotone itemkarakteristieke curve?

Wanneer we het hebben over uitspraken, beweringen, opinies e.d. vinden we alleen dan een monotone itemkarakteristiek bij extreme uitspraken; er zijn dan geen respondenten die bij een positief item 'nog positiever' zijn. Bijvoorbeeld bij (statistiek)-opgaven (dus: prestatie-items) komen monotone itemkarakteristieken vaak voor. Studenten hebben verschillende posities op het continuüm 'statistische kennis' :

de kans dat een student een opgave goed maakt, neemt monotoon toe naarmate de student meer naar rechts ligt op het continuüm, of eenvoudiger gezegd: de kans dat een student een opgave goed heeft, wordt groter naarmate hij beter is in statistiek.

Welke items hebben een Non-monotone itemkarakteristieke curve? Een dergelijke curve zal vooral worden aangetroffen worden bij items die gematigde uitspraken bevatten. Bijvoorbeeld:

- De OV-jaarkaart bevordert enigzins de mobiliteit van de studenten,
- Positieve discriminatie is een noodzakelijk kwaad.

Respondenten die sterk vóór of tegen zijn, zullen het niet eens zijn met deze items.

Wat is een schaal?

We begonnen deze paragraaf met de omschrijving van een schaal als een verzameling stimuli met betrekking tot één object. Thans kunnen we een uitgebreidere omschrijving geven.

Een schaal is een verzameling uitspraken over één bepaald object. Op de uitspraken wordt door de respondenten gereageerd. De respondenten worden verondersteld ten aanzien van het object in kwestie één (onderliggende) attitude te hebben. Een attitude is een latente variabele die de antwoorden verbindt op de items. De attitude stellen we ons voor als een continuüm dat loopt van zeer negatief tot zeer positief. Zowel respondenten als items nemen op dit continuüm een bepaalde positie in. De lijn die voor een item het verband aangeeft tussen de posities van de respondenten op het continuüm en de kans voor een respondent om 'ja' te zeggen bij dat item, heet de itemkarakteristieke curve. Er wordt onderscheidt gemaakt tussen monotone en non-monotone items.

De nu te behandelen schaaltechnieken verschillen in het gebruik van monotone en non-monotone items en in de manier waarop respondenten en items op het continuüm worden afgebeeld. De Thurstoneschaal maakt gebruik van non-monotone items, de Likertschaal en de Guttmanschaal maken gebruik van monotone items.

4 Likertschaal

De Likertschaal kunnen we als volgt omschrijven: Op grond van reacties van respondenten op stimuli worden getalswaarden aan de respondenten toegekend zonder dat de getalswaarden van de stimuli bekend zijn. De toegekende getalswaarden geven aan in welke mate elke respondent een bepaald kenmerk bezit. De stimuli zelf worden beschouwd als replicaties (herhalingen van elkaar) en hebben een monotone itemkarakteristiek, d.w.z. dat de items monotoon gerelateerd zijn aan één onderliggend attitude-continuüm: m.a.w. hoe gunstiger de attitude van de respondent, hoe hoger zijn score op het item. De items tezamen meten alléén het attribuut in kwestie, dus factoranalyse levert één (algemene) factor oplevert en de waarde van Cronbach's alfa zal hoog zijn. We zullen deze algemene, abstracte omschrijving via een voorbeeld nader uitwerken.

Als voorbeeld nemen we de onderzoeker met de hypothese dat slechte schoolprestaties van sommige kinderen terug te voeren zijn op de opvoedingssituatie in het ouderlijk gezin. De onderzoeker vindt het daarom nodig dat inzicht verkregen wordt in de attitude van de ouders ten aanzien van het opvoeden van kinderen. Deze attitude wordt geoperationaliseerd als: de mate waarin kinderen door hun ouders acceptierend en positief tegemoet worden getreden. Wat de onderzoeker dan nodig heeft, is een getalswaarde die hij aan een ouder kan toekennen en die de mate weerspiegelt waarin de ouder het kenmerk "acceptierend en positief optreden ten aanzien van kinderen" bezit. De onderzoeker kiest voor een attitudeschaal om de mening van de ouders over het opvoeden van kinderen te meten. Een voorbeeld is de door Itkin ontworpen schaal waaruit we zes voorbeelden geven³. De gehele schaal bestaat uit 26 items van het Likerttype.

³ Deze items zijn willekeurig gekozen uit de schaal van Itkin en zijn vrij vertaald.

1. Wanneer een kind minder snel met leren is dan de overigen, moeten zijn ouders hem aansporen door hem steeds te wijzen op het betere leren van de andere kinderen in het gezin.
 - a. zeer mee eens
 - b. mee eens
 - c. eens, noch oneens
 - d. niet mee eens
 - e. helemaal niet mee eens
2. Een ouder moet nooit aan een kind toegeven.
3. Kinderen moeten zo vroeg mogelijk in het leven leren om dingen zelf te doen.
4. Een gezin zou omwille van de kinderen uit een onprettige buurt moeten verhuizen, ook al zou het voor de vader betekenen dat hij langere reistijden voor zijn werk zou moeten maken.
5. Ouders moeten verzoeken aan hun kinderen in de basisschoolleeftijd of ouder altijd met redenen omkleden.
6. Als een kind een standje verdient, moet hij dat op dat moment krijgen of er nu wel of geen buitenstaanders bij aanwezig zijn.

Welke zijn nu de kenmerken van een vragenlijst als bovenstaand, dus wat zijn de kenmerken van een Likertschaal? Na een instructie over wat de antwoorden betekenen en over hoe de vragenlijst in te vullen, volgen een aantal uitspraken die een monotone itemkarakteristieke curve moeten hebben. Dus de items moeten zo geformuleerd zijn dat hoe hoger de positie van een respondent op het onderliggende attitude-continuüm, hoe meer hij het eens is met het item. Hieruit volgt dat elke uitspraak extreem geformuleerd moet zijn, dat wil zeggen extreem positief of extreem negatief, bijvoorbeeld:

- een ouder moet nooit aan een kind toegeven (negatief item).
- een ouder moet zo veel mogelijk tijd met zijn/haar kind doorbrengen (positief item).

Deze items zijn duidelijk interpreteerbaar en de respondent zal geen moeite hebben om aan te geven in hoeverre hij het met deze uitspraken eens is. Een voorbeeld van een minder duidelijk item is:

- hoe men ook over communisme denkt, in een aantal opzichten bevat het goede ideeën.

De moeilijkheid met deze uitspraak is dat een tegenstander van het communisme op dit item met "niet mee eens" kan antwoorden omdat hij het item te sterk vindt geformuleerd; hij kan van mening zijn dat communisme alleen maar foute ideeën bevat. Voorstanders van het communisme (communisten dus) kunnen het oneens zijn met dit item omdat zij het te zwak vinden geformuleerd. Dus respondenten met een verschillende attitude kunnen op dit item eenzelfde antwoord geven.

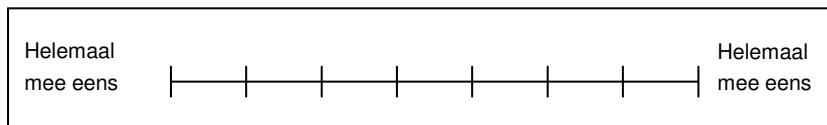
Daarom zijn dit soort items in een Likertschaal niet bruikbaar. Ook onbruikbaar zijn items die niet discrimineren omdat zij door iedereen worden onderschreven. Een item als "roken is slecht voor de gezondheid" zal door geen zinnig mens worden ontkent en dus discrimineert dit item niet

tussen personen met een positieve of negatieve attitude t.a.v. het roken. Behalve dat Likertitems extreem geformuleerd dienen te worden, moeten zij kort geformuleerd worden en dienen dubbele ontkenningen in de formulering te worden vermeden.

In een Likertschaal geven de personen op een continuüm (meestal een 5- of 7-puntsschaal) aan in welke mate zij het eens of oneens zijn met elk van de uitspraken die in de schaal zijn opgenomen. De responsecategorieën van een 5-puntsschaal kunnen de volgende zijn:

- zeer mee eens
- mee eens
- eens, noch oneens
- niet mee oneens
- helemaal niet mee eens

Het is ook mogelijk om deze 5 (of 7) responsecategorieën op een lijn uit te zetten en vervolgens laat men dan de persoon een kruisje zetten waarmee aangegeven wordt in hoeverre hij het eens/oneens is met de uitspraak.



Als men een grote groep respondenten verwacht die 'eens, noch oneens' zal aankruisen, kan men eventueel een 6-puntsschaal als onderstaande gebruiken:

- zeer mee eens
- mee eens
- eerder wel dan niet mee eens
- eerder niet dan wel mee eens
- oneens
- helemaal mee oneens

Hoe construeren we nu een Likertschaal?

De constructiestappen zijn de volgende:

Stap 1

Men verzamelt een groot aantal extreem geformuleerde uitspraken (15 of meer) die relevant zijn voor de betreffende attitude en die individuele verschillen in die attitude tot uitdrukking zullen brengen. Formuleer de ene helft van de items positief en de andere helft van de items negatief, omdat, wanneer er te veel items zijn die uitsluitend positief (of: uitsluitend negatief) geformuleerd zijn, er een neiging tot antwoordtendentie ontstaat (bv. een instemmingstendentie).

Opmerking

Men kan tijdens de analyse nagaan of de positief geformuleerde vragen hetzelfde meten als de negatief geformuleerde vragen. Als dat inderdaad het geval is, mag men verwachten dat de ge-

middelden van beide groepen items gelijk zijn aan elkaar. Zijn de gemiddelden duidelijk niet gelijk dan heeft men een probleem. De onderzoeker moet zich dan afvragen of hij niet twee begrippen heeft gemeten in plaats van één begrip. Men kan dan óf twee scores hanteren óf men kan één van de twee begrippen laten vallen .

Stap 2

Deze verzameling uitspraken legt men voor aan een groep personen die representatief is voor de populatie waar de schaal uiteindelijk voor is bedoeld (dus geen steekproef eerste jaarsstudenten in de psychologie die in het kader van hun testweek allerlei vragenlijsten moeten invullen en testjes moeten maken). De personen in deze proefgroep geven van elke uitspraak aan in hoeverre zij het er mee eens of oneens zijn met behulp van een 5- (6-, of een 7-)puntsschaal.

Stap 3

Aan elke responsecategorie wordt een score toegekend, zodanig dat een gunstige attitude de hoogste score krijgt. In de 6 voorbeelditems zullen bij de items 3, 4 en 5 de responsecategorie "zeer mee eens" de score 5 toegewezen krijgen (positieve items), maar bij de items 1, 2 en 6 (negatieve items) zal het de responsecategorie "helemaal mee oneens" zijn die de score 5 toegewezen krijgt. In schema:

	Score	Responsecategorie	Score	
Positieve Items	5	zeer mee eens	1	Negatieve Items
	4	mee eens	2	
	3	een, noch oneens	3	
	2	niet mee eens	4	
	1	helemaal mee oneens	5	

De score van een persoon krijgt men door zijn itemscores op te tellen. Dus als een Likertschaal uit 6 items bestaat met elk 5 responsecategorieën, dan is de minimumscore gelijk aan 6 en de maximumscore gelijk aan 30.

Stap 4

Na afname van de schaal aan de proefgroep worden de scores geanalyseerd teneinde vast te stellen welke items het beste discrimineren tussen personen met een hoge score (gunstige attitude) en personen met een lage score (ongunstige attitude).

De analyse begint met het berekenen van de coëfficiënt α .

Vervolgens berekent men van elk item de item-totaalcorrelatie (de r_{it} -waarde). Items met een hoge r_{it} -waarde discrimineren goed. Daarentegen worden items met een lage r_{it} -waarde verwijderd omdat dit soort items eerder tegenwerken dan dat zij bijdragen aan de totaalscore. Vaak worden items met een lage r_{it} -waarde ook verwijderd omdat dan een hoge totaalscore kan voorkomen ongeacht het resultaat op het item. Men kan onder andere lage r_{it} -waarden verwachten wanneer -ondanks betrachte zorgvuldigheid bij de constructie van de schaal- er toch (te) neutraal

geformuleerde items in de schaal zijn opgenomen. Men verwacht dan van deze items lage rit-waarden (deze items zullen laag met de totaalscore correleren) omdat degenen die op een (te) neutraal geformuleerd item "oneens" hebben aangekruist (zie voorbeelditem over communisme), zich over twee geheel verschillende delen van het continuüm zullen verspreiden en dus op een extreem geformuleerd item zeer verschillend zullen scoren. Na verwijdering van items met een lage rit-waarde berekent men opnieuw coëfficiënt α en men verwacht dat deze dan zal zijn gestegen.

NB: In paragraaf 5 geven we een uitgebreid voorbeeld hoe men door het verwijderen van items een hoge coëfficiënt alfa kan bereiken

De vraag doet zich vervolgens voor hoe hoog moet coëfficiënt α zijn, willen we van een acceptabele hoogte kunnen spreken? Een vuistregel is dat als het aantal items zich tussen de 15 en 20 bevindt, de coëfficiënt α dan hoger dan 0,80 moet zijn. Wordt deze hoogte niet meteen na de (proef)afname gehaald, dan kan men een laag correlerend item verwijderen om vervolgens de coëfficiënt α opnieuw te berekenen, enz. tot deze coëfficiënt wel boven de 0,80 uitkomt.

Stap 5

Soms wordt een frequentieverdeling van scores vereenvoudigd door deze op te delen in een kleiner aantal klassen (minimaal 2, meestal niet meer dan 5 of 6) met een zoveel mogelijk gelijke vulling per klasse. Dit in verband met een mogelijke multivariate analyse waarin het gewenst is om in tabellen met 4 of 5 ingangen toch nog redelijke celvullingen te houden. Een andere manier om een frequentieverdeling te vereenvoudigen is het streven naar een normale verdeling, bijvoorbeeld via een normaliserende transformatie naar staninescores.

5 Coëfficiënt α en item-totaalcorrelaties

In het algemeen streeft men naar een hoge coëfficiënt α . Men kan dit bereiken door veel items in een schaal op te nemen, ook al correleren die onderling laag. Men heeft dan weliswaar een schaal met een hoge betrouwbaarheid, maar men weet dan niet altijd goed wat deze schaal nu eigenlijk meet.

Een andere manier om een hoge coëfficiënt α te bereiken, is om items te verwijderen die laag met het totaalresultaat correleren. Na het verwijderen van items met lage r_{it} -waarden, berekent men opnieuw de coëfficiënt α en de r_{it} -waarden. Eventueel kan men deze gang van zaken herhalen tot coëfficiënt α niet meer stijgt.

We zullen een en ander toelichten aan de hand van een voorbeeld. Aan eerstejaars studenten opvoedkunde is in 2002 in het kader van onderwijsdoeleinden een vragenlijst over 'locus of control' afgenomen. Het begrip 'locus of control' kan het best omschreven worden door de twee polen van het begrip te omschrijven, nl. interne locus of control vs externe locus of control:

Interne locus of control

Sommige mensen verwachten dat vooral hun eigen gedrag (en eventueel ook kenmerken die zij zelf hebben) de uitkomsten van hun gedrag bepalen.

Externe locus of control

Anderen verwachten dat de uitkomsten (of meer algemeen de gebeurtenissen die zij in hun leven meemaken) vooral worden bepaald door omstandigheden buiten hen zelf, zoals toeval of de invloed van andere personen.

Hoewel de vragenlijst van 18 items door 98 studenten werd ingevuld, presenteren wij hier alleen de resultaten van de eerste 15 studenten op de eerste 10 items. De 10 items waarvan wij de resultaten hier zullen bespreken, luiden als volgt:

1. De meeste narigheden in je leven overkomen je zonder dat je er veel aan kunt doen.
2. Als je werkelijk goed je best doet, dan bereik je ook veel.
3. Het heeft geen zin om je recht te krijgen bij hogere ambtenaren, want die trekken zich van de gewone burger toch niets aan.
4. Hoe sterk iemand zich ook inspant, vaak wordt zijn waarde niet erkend.
5. Ik voel mij vaak een slachtoffer van de omstandigheden.
6. De werkelijke beslissingen worden genomen door een paar mensen die de macht hebben en de gewone burger kan daar niet zoveel aan doen.
7. Heel wat ongevallen zijn gewoon een kwestie van pech hebben.
8. Hoger komen in het leven heb je vrijwel zelf in handen.
9. Veel beslissingen zouden we net zo goed kunnen nemen door een munt op te gooien.
10. Je bent nu eenmaal een brokkenmaker of niet en daar is niet zo erg veel aan te veranderen.

Bij elke vraag waren 6 alternatieven:

- 1 = geheel oneens
- 2 = grotendeels oneens
- 3 = enigszins oneens
- 4 = enigszins eens
- 5 = grotendeels eens
- 6 = geheel mee eens

De bedoeling is om de vragen zo te scoren dat hoe hoger de somscore was hoe hoger de 'interne locus of control'. Zo opgevat, zijn de vragen 2 en 8 positief geformuleerd (hoe hoger hoe meer een interne locus of control) en de vragen 1, 3, 4, 5, 6, 7, 9 en 10 negatief (hoe hoger hoe meer een externe locus of control) geformuleerd. Dit betekent dat de negatief geformuleerde vragen omgescoord moeten worden, dus: itemscore 1 wordt 6; $2 \rightarrow 5$; $3 \rightarrow 4$; $4 \rightarrow 3$; $5 \rightarrow 2$ en $6 \rightarrow 1$. Na omscoring waren de resultaten van de 15 studenten op de 10 items als volgt:

tabel 1: Resultaten van 15 studenten op 10 locus of control-items

per- soon	items									
	1	2	3	4	5	6	7	8	9	10
1	3	4	3	3	3	3	1	4	1	6
2	5	5	4	6	5	5	3	4	5	6
3	4	3	5	6	6	5	2	5	6	3
4	4	5	4	5	5	3	4	4	5	5
5	2	5	5	5	4	5	5	4	6	5
6	6	4	6	6	6	6	4	4	6	4
7	2	5	4	4	5	5	3	3	3	5
8	4	5	3	3	4	3	3	4	5	5
9	2	4	4	3	4	4	4	4	4	5
10	3	6	5	5	4	4	5	6	5	4
11	3	6	6	3	5	3	3	6	6	6
12	3	4	5	5	4	5	3	5	6	5
13	1	5	5	5	6	5	3	6	4	5
14	4	5	6	4	4	4	4	4	6	5
15	3	5	5	5	5	4	3	4	5	5

We voeren de gegevens van tabel 1 in SPSS en laten vervolgens coëfficiënt α berekenen. Deze blijkt gelijk te zijn aan 0,65. De betrouwbaarheid is dus niet erg hoog. SPSS geeft ons ook de gemiddelde correlatie tussen de items. Deze blijkt te zijn: $r_{it} = 0,13$ en dat is aan de lage kant. Als we op de correlatiematrix van de items een principale componentenanalyse uitvoeren en we laten één component berekenen, dan verklaart deze component slechts 37% van de variantie. We kunnen dus niet zeggen dat er maar één 'onderliggende' factor is die de correlaties tussen de items kan verklaren.

Uit de gemiddelde item-item-correlatie en ook uit de correlaties zelf, kunnen we niet afleiden welke items in aanmerking komen om verwijderd te worden. SPSS geeft ook een tabel met de gecorrigeerde item-totaalcorrelaties⁴ alsmede de waarde van de α indien dat item wordt verwijderd.

tabel 2: Gecorrigeerde itemtotaalcorrelaties van de items uit tabel 1 en de waarde van coëfficiënt α indien dat item wordt verwijderd

item	itemtotaal- correlatie	α indien item verwij- derd
r1X	0,18	.66
r2X	0,08	.66
r3X	0,69	.55
r4X	0,56	.57
r5X	0,45	.66
r6X	0,37	.61
r7X	0,37	.61
r8X	0,18	.65
r9X	0,71	.51
r10X	-0,47	.74

We zien dat item 10 ("Je bent nu eenmaal een brokkenmaker of niet en daar is niet zo erg veel aan te veranderen.") negatief correleert met het totaalresultaat. Klaarblijkelijk meet dit item iets anders dan de overige items. Als we dat item verwijderen dan stijgt α van 0,65 tot 0,74. We verwijderen item 10 en berekenen opnieuw de coëfficiënt α .

Betrouwbaarheidsanalyse op 9 items (item 10 verwijderd)

Door weglating van item 10 stijgt coëfficiënt α van 0,65 tot 0,74. Ook de gemiddelde item-itemcorrelatie is gestegen en wel van 0,13 tot 0,23. De eerste principale component verklaart nog steeds slechts 37%.

⁴ zie verderop in deze paragraaf

tabel 3: Gecorrigeerde itemtotaalcorrelaties van de items 1 t/m 9 en de waarde van coëfficiënt α , indien dat item wordt verwijderd

item	itemtotaal-correlatie	α indien item verwijderd
r1X	0,19	.76
r2X	0,00	.76
r3X	0,70	.67
r4X	0,61	.67
r5X	0,50	.70
r6X	0,42	.71
r7X	0,38	.72
r8X	0,19	.74
r9X	0,73	.64

De item-totaalcorrelaties zijn over het geheel genomen gestegen, behalve de item-totaalcorrelatie van item 2. Als we item 2 verwijderen stijgt α van 0,74 tot 0,76; een geringe stijging. We besluiten desondanks om item 2 te verwijderen en we doorlopen de procedure opnieuw.

Betrouwbaarheidsanalyse op 8 items (item 2 en 10 verwijderd).

De coëfficiënt α stijgt van 0,74 tot 0,76 en de gemiddelde item-item-correlatie stijgt van 0,23 tot 0,29. De variantie die nu door de eerste component wordt verklaard is gestegen van 37% tot 41%.

tabel 4: Gecorrigeerde itemtotaalcorrelaties van de items (zonder item 2 en 10) en de waarde van coëfficiënt α indien dat item wordt verwijderd

	r _{ir}	α indien item verwijderd
r1X	0,24	.78
r3X	0,68	.70
r4X	0,67	.70
r5X	0,53	.73
r6X	0,49	.73
r7X	0,31	.76
r8X	0,15	.78
r9X	0,72	.68

Item 8 ("Hoger komen in het leven heb je vrijwel zelf in handen") is nu het item met de laagste item totaalcorrelatie, namelijk 0,15. We besluiten om dit item te verwijderen en we berekenen opnieuw coëfficiënt α .

Coëfficiënt α over 7 items (2, 8 en 10 zijn verwijderd)

De coëfficiënt α stijgt van 0,76 tot 0,78 en de gemiddelde item-itemcorrelatie stijgt van 0,29 tot 0,35. De variantie die nu door de eerste component wordt verklaard, stijgt van 41% tot 46%.

tabel 5: Gecorrigeerde itemtotaalcorrelaties van de items (zonder items 2, 8 en 10) en de waarde van coëfficiënt α indien dat item wordt verwijderd

	r_{ir}	α indien item verwijderd
r1X	.30	.80
r3X	.62	.74
r4X	.69	.72
r5X	.51	.76
r6X	.53	.75
r7X	.31	.79
r9X	.69	.71

Het item dat nu het laagst correleert met het totaalresultaat is item 1. We verwijderen dit item en berekenen nogmaals coëfficiënt α .

Coëfficiënt α over 6 items (items 1,2, 8 en 10 verwijderd)

De coëfficiënt α stijgt van 0,78 tot 0,80 en de gemiddelde item-itemcorrelatie stijgt van 0,35 tot 0,41. De variantie die nu door de eerste component wordt verklaard, stijgt van 46% tot 52%.

tabel 6: Gecorrigeerde itemtotaalcorrelaties van de items (zonder items 1,2, 8 en 10) en de waarde van coëfficiënt α indien dat item wordt verwijderd

	r_{ir}	α indien item verwijderd
r3X	.66	.75
r4X	.64	.75
r5X	.51	.78
r6X	.57	.77
r7X	.36	.81
r9X	.65	.75

Het item dat nu de laagste correlatie met het totaalresultaat heeft, is item 7. We verwijderen ook dit item en we berekenen weer coëfficiënt α .

Coëfficiënt α over 5 items (items 1,2, 7, 8 en 10 verwijderd)

De coëfficiënt α stijgt van 0,80 tot 0,81 en de gemiddelde item-itemcorrelatie stijgt van 0,41 tot 0,49. De variantie die nu door de eerste component wordt verklaard, stijgt van 52% tot 59%.

tabel 7: Gecorrigeerde itemtotaalcorrelaties van de items (zonder items 1,2, 7, 8 en 10) en de waarde van coëfficiënt α indien dat item wordt verwijderd

	r_{ir}	α indien item verwijderd
r3X	.63	.77
r4X	.68	.75
r5X	.60	.78
r6X	.60	.78
r9X	.57	.81

We zien dat alle r_{it} -waarden 0,57 of hoger zijn. Wanneer we nu nog weer een item (welk item dan ook) zouden verwijderen, zou de coëfficiënt α gelijk blijven of dalen. Het heeft dus geen zin om nog meer items te verwijderen. We vatten de voorgaande resultaten samen in onderstaande tabel.

tabel 8: Samenvatting van analyse van α en item-totaalcorrelaties

		gemiddelde correlatie tussen de items	1e component ver- klaart	gemiddelde gecorrigeerde item-totaal correlatie	item met de laagste r-ir-corre- latie
k	α				
10	0,65	0,13	37%	0,31	10: -0,47
9	0,74	0,23	37%	0,41	2: 0,00
8	0,76	0,29	41%	0,47	8: 0,15
7	0,78	0,35	46%	0,52	7: 0,30
6	0,80	0,41	52%	0,57	1: 0,36
5	0,81	0,49	59%	0,62	Als we nog een item weg- laten,daalt α

We zijn begonnen met 10 items met een α gelijk aan 0,65 en we zijn geëindigd met 5 items met een α gelijk aan 0,81. De eerste component verklaart nu 59% van de variantie. Een tweede component zou een eigenwaarde van 0,99 hebben.

Er is overigens geen duidelijk criterium dat aangeeft wanneer we moeten stoppen met het verwijderen van items. Is een stijging van α van 0,74 tot 0,75 nog de moeite waard om daarvoor een item te verwijderen? Wanneer men toch al te veel items heeft, dan wellicht wel. Maar ook inhoudelijke argumenten moeten een rol spelen: schaadt het weglaten van een item de inhoudsvaliditeit? Neem het volgende voorbeeld (J.P.v.d. Sande, 1984, bl.97). Van 5 kinderen zijn via observatie agressieve gedragingen geturfd. De fictieve gegevens zijn als volgt:

tabel 9: Aantallen agressieve gedragingen van 5 kinderen

	schel- den	uit- lachen	slaan	knij- pen	schop- pen	totaal
kind 1	4	10	1	0	0	15
kind 2	0	0	0	2	0	2
kind 3	8	17	0	4	1	30
kind 4	10	16	4	0	2	32
kind 5	6	15	3	0	0	24

De item-totaalcorrelaties staan hieronder.

tabel 10: Itemtotaalcorrelaties

	Item- totaalcor- relaties
schelden	0,99
uitlachen	0,97
slaan	0,56
knijpen	0,00
schoppen	0,73

'Knijpen' correleert in het geheel niet met het totaalresultaat en zou dus verwijderd moeten worden. Echter, het doet wat merkwaardig aan om in een maat voor 'agressie' op louter statistische gronden een variabele als 'knijpen' weg te laten. De coëfficiënt α zal ongetwijfeld stijgen, maar de betekenis van het begrip 'agressie' verandert dan.

De procedure voor het verwijderen van items heeft weliswaar als gevolg dat er een homogenere schaal ontstaat, maar die procedure kan dus ook tot gevolg hebben dat de betekenis van de schaal verandert. Het kan zelfs zijn dat men een ander begrip meet dan welk men voor ogen had. Wanneer items met een lage rit-waarde betrekking hebben op een specifiek inhoudelijk terrein, dan verandert na verwijdering van items het begrip. Men zou dan eigenlijk het item of de items moeten herschrijven en de schaal opnieuw moeten afnemen aan een proefgroep, enzovoort. Het voorgaande illustreert het gevaar van het gedachteloos toepassen van 'automatische' verwijderingsprogramma's. Men dient dus ook aandacht te besteden aan de inhoudelijke aspecten.

In het algemeen verwachten we dus positieve correlaties tussen de afzonderlijke items en het totaalresultaat. Hoe hoger de r -it-waarden zijn hoe hoger is in het algemeen de coëfficiënt α . Bij dichotome items (dus items die met 0 of 1 gescoord worden) kan worden aangetoond dat het verband tussen de coëfficiënt α en de r -it-waarden gelijk is aan

$$\alpha = \frac{k}{k-1} \left[1 - \frac{\sum p q}{[\sum r_{it} \sqrt{p q}]^2} \right]$$

We hebben in het voorgaande reeds gesproken over de gecorrigeerde itemtotaalcorrelaties. Hiermee wordt het volgende bedoeld. In het totaalresultaat zit ook het resultaat van het betreffende item verwerkt. Dus men correleert bij het berekenen van een item-totaalcorrelatie een item deels met zichzelf en dat maakt dat de item-totaalcorrelatie iets hoger is dan hij zou moeten zijn. Het is zuiverder om uit het totaalresultaat het resultaat van het afzonderlijke item te verwijderen en dan de correlatie te berekenen tussen het item en de totaalscore (minus het resultaat op dat item). Deze correlaties heten de item-restcorrelaties (r -ir- waarden) ofwel de gecorrigeerde item-

totaalcorrelaties. Voor de voorgaande datamatrix van 10 items zijn de r-it- en de r-ir-waarden gelijk aan:

tabel 11: Verschil tussen r-it-waarden en r-ir-waarden

rit	rir	verschil
0,42	0,18	0,24
0,23	0,08	0,15
0,79	0,69	0,10
0,70	0,56	0,14
0,59	0,45	0,14
0,53	0,37	0,16
0,54	0,37	0,17
0,35	0,18	0,17
0,83	0,71	0,12
-0,34	-0,47	0,13

Over het algemeen zijn de gecorrigeerde item-totaalcorrelaties dus lager dan de (ongecorrigeerde) item-totaalcorrelaties. Maar als het aantal items groot is, dan maakt het niet zoveel verschil. SPSS geeft de gecorrigeerde item-totaalcorrelaties. Voor de volgorde waarin items verwijderd worden, maakt het niet veel uit of we nu met r-it-waarden of met r-ir-waarden werken.

6 Voorbeeld schaalanalyse

We behandelen het volgende voorbeeld. Een schaal -- Rosenberg Self-Esteem Scale-- is afgenomen aan 1000 highschool leerlingen van 25 verschillende scholen in de staat Pennsylvania. Het voorbeeld is overgenomen uit: Mclver, J.P & Carmines, E.G. (1981). Unidimensional scaling. London, Sage Publications, nr.24.

De items waren:

1. I feel that I have a number of good qualities.
2. I wish I good have more respect for myself.
3. I feel I'm a person of worth, at least on an equal plane with others.
4. I feel I do not have much to be proud of.
5. I take a positive attitude toward myself.
6. I certainly feel useless at times.
7. All in all, I am inclined to feel that I am a failure.
8. I am able to do things as well as most other people.
9. At times I think I am no good at all.
10. On the whole, I am satisfied with myself.

Bij elk item waren de volgende alternatieven gegeven:

- ☐ almost always true
- ☐ often true
- ☐ sometimes true
- ☐ seldom true
- ☐ never true

De correlatiematrix wordt vermeld in tabel 12.

tabel 12: Correlatiematrix van 10 variabelen

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
V1	1.00	0.16	0.46	0.50	0.41	0.23	0.33	0.39	0.32	0.40
V2	0.16	1.00	0.22	0.29	0.29	0.28	0.22	0.13	0.24	0.21
V3	0.46	0.22	1.00	0.30	0.42	0.24	0.36	0.42	0.33	0.43
V4	0.50	0.29	0.30	1.00	0.35	0.32	0.45	0.25	0.37	0.29
V5	0.41	0.29	0.42	0.35	1.00	0.37	0.36	0.35	0.39	0.49
V6	0.23	0.28	0.24	0.32	0.37	1.00	0.38	0.21	0.46	0.28
V7	0.33	0.22	0.36	0.45	0.36	0.38	1.00	0.31	0.51	0.41
V8	0.39	0.13	0.42	0.25	0.35	0.21	0.31	1,00	0.24	0.38
V9	0.32	0.24	0.33	0.37	0.39	0.46	0.51	0,24	1,00	0.38
V10	0.40	0.21	0.43	0.29	0.49	0.28	0.41	0,38	0,38	1,00

De coëfficiënt alpha was: 0.84. De item-totaalcorrelaties worden in onderstaande tabel vermeld:

tabel 13: Item-totaalcorrelaties

Varia- bele	r_{it}	Varia- bele	r_{it}
V1	0.66	V6	0.60
V2	0.48	V7	0.68
V3	0.66	V8	0.58
V4	0.65	V9	0.67
V5	0.70	V10	0.68

Conclusie

Er is niet direct een reden om items te verwijderen. Maar is er sprake van een homogene schaal?
Vandaar dat we nu eerst een PCA uitvoeren waarvan de resultaten als volgt zijn.

tabel 14: Resultaten van de principale componentenanalyse op de correlatiematrix in tabel 12, één-factor-oplossing

	ladingen	communaliteiten
V1	0.67	0,45
V2	0.44	0,19
V3	0.67	0,44
V4	0.65	0,42
V5	0.71	0,50
V6	0.59	0,34
V7	0.69	0,48
V8	0.58	0,33
V9	0.68	0,46
V10	0.69	0,47

De factorladingen zijn allemaal groter dan 0,50 behalve die van item 2. Deze één-factor-oplossing verklaart 41% van de variantie. Wanneer we een twee-factor-oplossing proberen, krijgen we twee factoren met eigenwaarden groter dan 1. De resultaten zijn als volgt:

tabel 15: Resultaten van de principale componentenanalyse op de correlatiematrix in tabel 11, twee-factor oplossing

	Ongeroteerde componentenmatrix		
	I	II	h^2
V1	0.67	-0.34	0,47
V2	0.44	0.44	0,39
V3	0.67	-0.35	0,57
V4	0.65	0.14	0,44
V5	0.71	-0.06	0,50
V6	0.59	0.48	0,57
V7	0.69	0.19	0,51
V8	0.58	-0.47	0,55
V9	0.68	0.33	0,56
V10	0.69	-0.21	0,51

Deze componentenmatrix zie er na varimaxrotatie als volgt uit:

tabel 16: Resultaten van de principale componentenanalyse op de correlatiematrix in tabel 11, twee-factor oplossing

	Geroteerde componentenmatrix	
	I	II
V1	0.73	0.20
V2	0.03	<u>0.62</u>
V3	0.73	0.18
V4	0.39	<u>0.54</u>
V5	0.56	0.43
V6	0.11	<u>0.75</u>
V7	0.39	<u>0.60</u>
V8	0.74	0.04
V9	0.29	<u>0.69</u>
V10	0.65	0.30

De geroteerde twee-factor-oplossing identificeert twee clusters van items: het eerste cluster bestaat uit de items 1, 3, 5, 8 en 10 (vetgedrukt in bovenstaande tabel); het tweede cluster bestaat uit de items 2, 4, 6, 7 en 9 (onderstreept). We zouden nu 2 schalen kunnen construeren:

schaal a Items 1, 3, 5, 8 en 10; Coefficient alpha: 0.78

Variabele	rit
V1	0.73
V3	0.75
V5	0.73
V8	0.70
V10	0.74

schaal b Items 2, 4, 6, 7 en 9; Coefficient alpha: 0.73

Variabele	rit
V2	0.58
V4	0.70
V6	0.71
V7	0.74
V9	0.74

Hoewel het voorgaande mathematisch correct is, kunnen we toch geen inhoudelijke interpretatie aan dit patroon toekennen. Voor verdere aanwijzingen keren we daarom terug naar de ongeroteerde oplossing. Factor 1 is in essentie gelijk aan de één-factor-oplossing, terwijl de 2e factor vijf negatieve ladingen laat zien. Dit zijn precies de items die in de geroteerde oplossing de hoogste lading hadden op factor 1. Dit patroon van negatieve en positieve ladingen op de tweede factor van de ongeroteerde factoroplossing correspondeert precies met het patroon van positief en negatief geformuleerde items. Dat wil zeggen we zien dat de items 1, 3, 5, 8 en 10 aan de respondent vragen om 'goede kwaliteiten' te beoordelen, terwijl de rest van de items vragen om 'slechte kwaliteiten' te beoordelen. Dus we interpreteren de tweede factor als een 'response set factor', niet als een inhoudelijk verschillende dimensie. We kunnen nu besluiten om alsnog slechts één totaalschaalscore te hanteren of om alleen subschaal a of subschaal b te gaan gebruiken.

Tot besluit volgen puntsgewijs enkele opmerkingen.

- 1) In voorgaande twee paragrafen is de betrouwbaarheidsanalyse behandeld. Betrouwbaarheid is nodig maar niet voldoende, het gaat uiteindelijk om de validiteit van de schaal: wat betekent de schaal en met welke andere variabelen is de schaal gecorreleerd? Voor de beantwoording van deze vragen is empirisch onderzoek nodig.
- 2) De procedure om met behulp van r_{it} -waarden items te verwijderen, heeft weliswaar als gevolg dat er een homogene schaal ontstaat, maar die procedure kan ook tot gevolg hebben dat de betekenis van de schaal verandert. Het kan zelfs zijn dat men een andere attitude gaat meten dan die men voor ogen had. Wanneer items met een lage r_{it} -waarde betrekking hebben op een specifiek inhoudelijk terrein, dan verandert na verwijdering van items de attitude. Men zou dan eigenlijk het item of de items moeten herschrijven en de schaal opnieuw moeten afnemen aan

een proefgroep, enz. Het voorgaande illustreert het gevaar van het gedachteloos toepassen van 'automatische' verwijderingsprogramma's. Men dient ook aandacht te besteden aan de inhoudelijke aspecten.

- 3) Behalve de analyse met behulp van r_{it} -waarden, kan men dus ook principale componentenanalyse (PCA) gebruiken.
Men voert op de correlatiematrix van de items een PCA uit waarbij men één component laat berekenen. Men verwacht dan uitsluitend hoge ladingen. Items met lage ladingen kan men verwijderen. Men berekent zowel voor als na de PCA de coëfficiënt α . Deze methode (PCA) heeft als bijkomend voordeel dat men een soort toets op de unidimensionaliteit uitvoert, men verwacht namelijk één algemene factor die veel van de variantie verklaart. Een grote algemene factor impliceert wel een hoge coëfficiënt α , maar een hoge coëfficiënt α impliceert niet per se een algemene factor. Zou men overigens niet één, maar twee factoren vinden dan kan men overwegen om twee schalen te construeren of om verder te gaan met items van één van de twee factoren en de items van de andere factor te verwijderen.
- 4) Soms is men van mening dat het ene item belangrijker is dan de andere items en dat item zou men dan zwaarder willen wegen. Voor dat wegen zijn verschillende methoden. Echter als de schaal homogeen is, heeft wegen weinig invloed. Als de schaal niet homogeen is dan is nader onderzoek naar de structuur van de schaal belangrijker. Dus in het algemeen geldt: niet wegen.

We hebben in het voorgaande de schaalconstructie volgens de Likertmethode behandeld. De Likertschaal is het meest voorkomende schaaltype en is ook het meest eenvoudig om te construeren. De Likertschaal kent scores toe aan personen en niet aan stimuli (items). Een schaaltechniek die schaalwaarden toekent aan items is de Thurstoneschaal.

7 Thurstoneschaal (methode van gelijkschijnende intervallen)

Thurstone heeft meerdere schaaltechnieken ontworpen; in deze paragraaf behandelen wij de Thurstoneschaal volgens de methode van gelijkschijnende intervallen. Voor een behandeling van de overige Thurstoneschalen verwijzen wij naar het boek Schaaltechnieken van Swanborn (1988).

Voor de constructie van een Thurstoneschaal moeten beoordelaars stimuli beoordelen op een continuüm lopend van 'zeer negatief' (ongunstig, etc.) tot 'zeer positief' (gunstig, etc.). Variaties in de reacties op de stimuli door de beoordelaars worden toegewezen aan verschillen tussen deze stimuli. Het gaat er om de stimuli te schalen en daaraan getalswaarden toe te kennen. De beoordelaars worden gezien als replicaties.

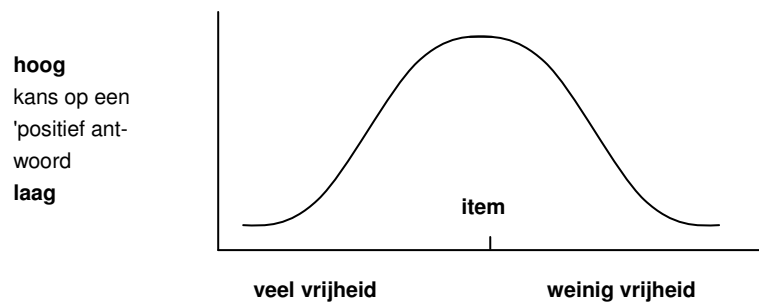
Deze algemene beschrijving vereist een nadere toelichting.

Een aantal beoordelaars (dat aantal kan variëren van 25 tot 100) krijgt bijvoorbeeld een groot aantal uitspraken voorgelegd over opvoeding met de vraag om deze uitspraken te beoordelen met betrekking tot de 'mate van vrijheid die ouders hun kinderen toestaan'. Deze uitspraken (bijv. een ouder moet, wanneer de kinderen ruzie maken, soms tussenbeide komen) representeren de stimuli (ofwel: de items). De veronderstelling is dat elke uitspraak geplaatst kan worden op het continuüm 'mate van toegestane vrijheid' lopend van bijvoorbeeld 1 (=zeer veel toegestane vrijheid) tot en met 11 (= zeer weinig toegestane vrijheid). Volgens Thurstone zullen de beoordelaars de categorieën 1 t/m 11 opvatten als intervallen die ongeveer een gelijke omvang hebben (vandaar de naam van de methode).

De beoordelaars moeten dus de uitspraken classificeren in 11 categorieën (dat kan ook een ander aantal zijn). De beoordelaar moet -los van zijn eigen opvatting- aangeven in hoeverre een uitspraak duidt op veel of weinig toegestane vrijheid. Als men als beoordelaar van mening is dat een bepaalde uitspraak zeer veel vrijheid aan een kind toekent, moet een 1 aan deze uitspraak worden toegekend, etc. Dus een belangrijke aanname in deze schaaltechniek is dat de eigen mening van de beoordelaar geen rol speelt bij hun oordeel.

De Thurstoneschaal volgens gelijkschijnende intervallen gaat uit van non-monotone items. Dus wanneer een uitspraak als 'een ouder moet, wanneer de kinderen ruzie maken, soms tussenbeide komen' bijvoorbeeld een schaalwaarde 6 heeft, verwacht men dat zowel respondenten die hun kinderen heel veel vrijheid toestaan als respondenten die hun kinderen weinig of geen vrijheid toestaan, het met deze uitspraak oneens zijn. De ene groep omdat zij de uitspraak te restrictief vinden en de andere groep omdat zij de uitspraak te weinig restrictief vinden. De kans op een positieve reactie op dit item neemt af naarmate de attitude van de ouder verder van de schaalwaarde 6 afligt. De item-karakteristieke curve zal er bij Thurstone-items als volgt uitzien.

Figuur 4: Karakteristiek curve van een Thurstone-item



Nadat de beoordelaars de items hebben beoordeeld op de schaal van 1 tot en met 11, berekent men per item het gemiddelde en de standaardafwijking. Het gemiddelde oordeel over een bepaald item representeert de schaalwaarde van dat item.

Vaak wordt de mediaan en de interkwartielafstand berekend. De schaalwaarden zullen dan wat verschillend zijn maar de volgorde van de schaalwaarden zal nauwelijks beïnvloed worden. De reordering bij de selectie van de items blijft overigens hetzelfde.

Selectie van items

Hoe komt men nu tot de selectie van de items? Want meestal laat men veel meer items beoordelen dan het aantal dat men in de uiteindelijke (attitude-)schaal wil opnemen. De vuistregels zijn de volgende:

- Men kiest die items waarvan de schaalwaarden over het gehele continuüm verspreid liggen;
- Wanneer twee items gelijke gemiddelden hebben, kiest men dat item met de kleinste standaarddeviatie. Dat item is het minst ambigu; de beoordelaars zijn het daar meer over eens dan over het item met de grotere standaarddeviatie.
- De items worden zo gekozen dat zij zo veel mogelijk op gelijke afstanden van elkaar liggen.

Wanneer men op deze wijze de uiteindelijke schaal heeft samengesteld, legt men deze voor aan respondenten die van elk item aangeven of zij het er wel of niet mee eens zijn. Elk item heeft dus slechts twee antwoordcategorieën: eens of oneens (voor of tegen). De veronderstelling is dat de respondent alleen instemt met die items die dezelfde positie op het continuüm innemen als hijzelf. Het gemiddelde van de schaalwaarden van de gekozen items (dus de items waarmee de respondent het eens is) representeert de attitude van de respondent in kwestie.

Wanneer de schaal inderdaad slechts één attitude meet, mag men verwachten dat een respondent alleen die items aanstreept met onmiddellijk op elkaar volgende schaalwaarden. Zeer gespreide scores kunnen er op duiden dat de respondent geen duidelijke mening heeft over het object in kwestie of dat zijn attitude niet is georganiseerd volgens de dimensie die door de schaal verondersteld wordt.

Het voordeel van een Thurstoneschaal is dat zonder veel extra werk uit de eerste itempool een parallelschaal ontwikkeld kan worden hetgeen van belang kan zijn bij het bestuderen van attitude-

veranderingen. Het nadeel van een Thurstoneschaal is dat het veel werk is om een dergelijke schaal te construeren. Het construeren van een Likertschaal gaat veel sneller.

Hoewel de absolute waarde van de schaalwaarden van onderzoek tot onderzoek zal verschillen, is de volgorde van de schaalwaarden in het algemeen wél hetzelfde.

In tegenstelling tot de aanname dat de eigen mening van de beoordelaar geen rol speelt bij de beoordeling, blijkt uit de literatuur dat beoordelaars die sterk betrokken zijn bij het object in kwestie en die dus meestal een extreme houding hebben, de neiging hebben om de items in een klein aantal categorieën te plaatsen. Een vuistregel die in de praktijk wel wordt toegepast is dat wanneer een beoordelaar meer dan een kwart van de items in één categorie plaatst, deze beoordelaar 'verwijderd' wordt.

Voorbeeld

In 1929 lieten Thurstone & Chave lieten 300 beoordelaars 130 uitspraken over de kerk indelen in 11 categorieën. Van deze schaal worden hieronder 10 voorbeelditems weergegeven⁵.

tabel 17: Voorbeeld van een Thurstoneschaal

Scale Value	Item
1,2	I believe the church is a powerful agency for promoting both individual and social righteousness.
2,2	I like to go to church for I get something worthwhile to think about and it keeps my mind filled with right thoughts.
3,3	I enjoy my church because there is a spirit of friendliness there.
4,5	I believe in what the church teaches but with mental reservations.
6,7	I believe in sincerity and goodness without any church ceremonies.
7,5	I think too much money is being spent on the church for the benefit that is being derived.
9,2	I think the church seeks to impose a lot of worn-out dogmas and medieval superstitions.
10,4	The church represents shallowness, hypocrisy, and prejudice.
11,0	I think the church is a parasite on society.

Voorbeeld

Een onderzoeker wil een attitude-schaal construeren volgens Thurstone's methode van gelijke intervallen. Hij wil uiteindelijk een schaal van 5 items overhouden. De onderzoeker verzamelt 15 uitspraken over de opvattingen van ouders over de vrijheid die aan kinderen toegestaan wordt.

De items luiden als volgt:

1. Men moet nooit van een kind verlangen dat het "alstublieft" zegt.
2. Een kleuter moet een zekere mate van vrijheid toegestaan worden om zijn eigen beslissingen te nemen en om daarvan de gevolgen te dragen.
3. Een kind moet in de keuze van zijn/haar vriendjes volledig vrij gelaten worden.
4. Het is zeer gewenst dat een kind leert dat het niet altijd zijn zin kan krijgen.
5. Een kind moet niet toegestaan worden om zijn speelgoed kapot te maken.
6. Een kind moet gedwongen worden om te gehoorzamen wanneer het niet onmiddellijk doet wat hem gezegd wordt.
7. Een kind moet nooit toegestaan worden om openlijk met zijn ouders van mening te verschillen.
8. Een ouder moet van een kind verlangen dat het eerst toestemming vraagt voordat een kind aan activiteiten begint die nieuw of vreemd zijn.
9. Een kind moet nooit gedwongen worden iets te doen tegen zijn zin.
10. Wanneer een ouder aan een kind bepaalde beperkingen oplegt, hoeft de ouder alleen dan de redenen op te geven wanneer de ouder van mening is dat het kind die redenen kan begrijpen.
11. Kinderen krijgen te veel vrijheid.
12. Behalve in gevaarlijke situaties mag men van een kind nooit verlangen dat het gehoorzaamt zonder dat er een goede reden voor gegeven wordt.
13. Wanneer een kleuter zo ver is dat hij zelf aangeeft welke kleren hij wel en welke kleren hij niet wil dragen, moet hem dat toegestaan worden.
14. Een kleuter moet nooit toegestaan worden om zijn zin door te drijven.
15. Een kind moet leren om de wensen van zijn ouders te respecteren.

Deze 15 items worden aan 25 respondent-beoordelaars voorgelegd met het verzoek om elk item te beoordelen op een schaal van

1 = heel veel toegestane vrijheid
(...)
11 = heel weinig toegestane vrijheid

Met nadruk wordt de respondent-beoordelaars gezegd dat het niet gaat om hun mening over de hoeveelheid vrijheid die men aan kinderen moet geven, maar dat het gaat om een beoordeling van de hoeveelheid vrijheid die in elk item tot uitdrukking wordt gebracht. De resultaten van de beoordelaars zijn als volgt:

⁵ Overgenomen uit: Hoyle a.o.2002

tabel 18: Resultaten van 25 respondent beoordelaars van 15 items op een schaal van 1-11

Be o	it_1	it_2	it-3	it-4	it-5	it-6	it-7	it-8	it-9	it-10	it-11	it-12	it-13	it-14	it-15
1	1	3	1	7	7	11	10	6	2	7	9	2	1	11	8
2	1	4	2	5	6	11	9	6	1	5	8	3	5	10	9
3	1	5	3	4	9	10	11	4	3	4	6	4	7	9	6
4	3	4	2	7	5	10	7	8	1	9	8	3	2	11	9
5	1	2	3	3	6	11	11	8	1	7	9	4	1	11	9
6	2	3	2	8	7	10	7	5	1	5	9	3	5	10	8
7	1	6	3	5	9	11	11	9	2	9	9	4	3	8	10
8	4	7	3	6	5	11	9	5	3	4	9	4	5	11	8
9	2	5	2	6	9	10	10	6	1	7	8	3	8	11	9
10	1	4	2	5	6	9	11	4	2	3	8	3	4	11	9
11	1	6	3	4	8	11	9	8	1	9	7	5	3	9	8
12	1	4	5	6	6	10	8	5	1	5	9	5	6	10	7
13	3	5	1	7	9	10	7	6	2	5	7	2	6	8	9
14	1	4	3	5	6	11	11	7	1	8	8	4	3	11	8
15	2	6	2	5	7	10	11	6	2	6	9	3	6	9	7
16	1	6	3	6	9	11	8	8	1	8	10	4	4	8	9
17	1	5	3	4	10	10	9	7	1	6	9	4	2	11	10
18	3	3	2	7	9	11	10	5	3	6	7	3	8	11	7
19	1	2	1	5	7	10	11	6	1	5	11	3	3	9	8
20	1	5	3	4	9	11	8	7	1	8	8	4	7	10	11
21	1	4	3	5	6	10	10	6	1	6	9	4	2	11	9
22	1	5	4	6	8	11	11	8	2	9	7	5	5	11	8
23	3	5	2	7	9	11	9	9	1	8	8	3	6	10	7
24	1	5	2	4	6	10	11	7	2	6	11	3	6	9	11
25	1	4	2	6	10	10	8	7	1	8	9	2	4	10	9

Bovenstaande tabel moet men als volgt lezen: respondent-beoordelaar 18 heeft aan uitspraak 9 (Een kind moet nooit gedwongen worden iets te doen tegen zijn zin) een 3 toegekend terwijl beoordelaar 19 aan diezelfde uitspraak een 1 toekende, enz.

Er zijn 5 beoordelaars die een kwart van de uitspraken (dat zijn 4 uitspraken) in één categorie hebben geplaatst:

- beoordelaar 3 heeft de uitspraken 4, 8 10 en 12 met een 4 beoordeeld;
- beoordelaar 7 heeft de uitspraken 5, 8 10 en 11 met een 9 beoordeeld;
- beoordelaar 12 heeft de uitspraken 3, 8 10 en 12 met een 5 beoordeeld;
- beoordelaar 16 heeft de uitspraken 7, 8, 10 en 14 met een 8 beoordeeld;
- beoordelaar 18 heeft de uitspraken 1, 2,9 en 12 met een 3 beoordeeld.

Wanneer dit een echt onderzoek zou zijn geweest, had de onderzoeker ernstig overwogen om de resultaten van deze beoordelaars niet mee te laten tellen. Overigens zou men in een echt onderzoek met meer beoordelaars en vooral met meer items hebben gewerkt.

We berekenen nu van elke uitspraak de gemiddelde beoordeling (= schaalwaarde) en de standaarddeviatie. In onderstaande tabel staan de resultaten die in volgorde van het gemiddelde worden weergegeven.

tabel 19: Gemiddelde beoordeling per uitspraak en de standaarddeviaties

itemnr	Gem.	St.Dev.
9	1.52	0.71
1	1.56	0.92
3	2.48	0.92
12	3.48	0.87
2	4.48	1.26
13	4.48	2.06
4	5.48	1.26
10	6.52	1.76
8	6.52	1.42
5	7.52	1.58
11	8.48	1.19
15	8.52	1.23
7	9.48	1.45
14	10.00	1.08
6	10.44	0.58

Hoe komen we nu tot de selectie van de 5 items die in de schaal moeten worden opgenomen?

Selectiecriteria voor de Thurstone-items

Hiervoor werden een aantal vuistregels genoemd om die items te selecteren die in de schaal opgenomen worden:

- schaalwaarden moeten over het gehele continuüm verspreid liggen;
De schaalwaarden moeten dan in de buurt liggen van de volgende waarden:
2,5; 4,5; 6,5; 8,5; 10,5.
- bij (praktisch) gelijke gemiddelden kiest men dat item met de kleinste standaarddeviatie;
- de schaalwaarden van de items moeten op zoveel mogelijk gelijke afstanden van elkaar liggen.

We komen dan tot de volgende items:

item 3

item 2 (Item 13 heeft hetzelfde gemiddelde maar een grotere sd).

item 8 (item 10 heeft hetzelfde gemiddelde maar een grotere sd)

item 11 (item 15 heeft hetzelfde gemiddelde en sd)

item 6

In de tabel zijn deze items grijs gekleurd. Deze manier van items selecteren vereist een 'timmermansoog' en is dus enigszins subjectief. Objectievere procedures geven echter niet zulke geheel andere resultaten.

We hebben nu de volgende attitudeschaal bestaande uit vijf items:

item 3 gem = 2,48

Een kind moet in de keuze van zijn/haar vriendjes volledig vrij gelaten worden.

item 2 gem = 4,48

Een kleuter moet een zekere mate van vrijheid toegestaan worden om zijn eigen beslissingen te nemen en om daarvan de gevolgen te dragen.

item 8 gem = 6,52

Een ouder moet van een kind verlangen dat het eerst toestemming vraagt voordat een kind aan activiteiten begint die nieuw of vreemd zijn.

item 11 gem = 8,48

Kinderen krijgen te veel vrijheid.

item 6 gem = 10,44

Een kind moet gedwongen worden om te gehoorzamen wanneer het niet onmiddellijk doet wat hem gezegd wordt.

Deze schaal presenteren we aan respondenten waarvan we willen weten wat hun houding is over de hoeveelheid vrijheid die men aan kinderen moet toestaan. De antwoorden van een (fictieve) respondent zijn bijvoorbeeld als volgt:

tabel 20: Antwoorden van een respondent

item	Antwoord respondent
2	oneens
3	oneens
6	eens
8	oneens
11	eens

De respondent is het eens met de items 6 en 11 waarvan de schaalwaarden respectievelijk zijn 10,44 en 8,48. De score op de attitude-schaal van de respondent is nu gelijk aan $(10,44 + 8,48) / 2 = 9,46$. Deze respondent is dus van mening men aan kinderen niet veel vrijheid moet toestaan.

8 Guttmanschaal (methode voor cumulatieve schalen)

Volgens Guttman kan men op grond van de resultaten van een Likertschaal of een Thurstoneschaal niet eenduidig vaststellen of een verzameling items op één onderliggend continuüm kunnen worden afgebeeld. Pas wanneer men uit de totaalscore (= somscore van de afzonderlijke items) de scores op de afzonderlijke items kan afleiden kan men volgens Guttman stellen dat de items op één continuüm kunnen worden afgebeeld. De Guttmanschaal is een procedure om zowel items als respondenten op één onderliggend continuüm af te beelden. Nadat de gegevens verzameld zijn, wordt getoetst of de data aan het Guttmanmodel voldoen. Eventueel worden er items verwijderd, waarna opnieuw getoetst wordt of aan het Guttmanmodel is voldaan.

Een Guttmanschaal werkt met dichotome antwoorden op items die verondersteld worden een monotone itemkarakteristiek te hebben. De Guttmanschaal dus gebaseerd op monotone items; elk item is duidelijk gunstig of ongunstig ten aanzien van een bepaald object. Wanneer men van een Likertschaal wil nagaan of zij ook voldoen aan een Guttmanschaal, moeten de antwoorden eerst gedichotomiseerd worden. Een Guttmanschaal bestaat dus uit items waarvan de respondent aangeeft of hij het er mee eens is of niet.

Voorbeeld Guttmanitems

Een voorbeeld uit een prestatietoets kan het voorgaande verduidelijken. Stel dat een wiskundetoets uit drie items bestaat:

- item a): eenvoudige optelsom (bijvoorbeeld 3+4),
- item b): een deelsom bestaande uit twee getallen van elk drie cijfers (bijvoorbeeld 875 : 321)
- item c): een opgave uit de differentiaalrekening (bijvoorbeeld voor welke waarde van X bereikt $y = 2X^2 + 3X + 4$ een maximum of een minimum?).

We verwachten nu dat iedereen die in staat is om opgave b) te maken ook in staat is om opgave a) te maken. Voorts verwachten we dat iedereen die opgave c) goed heeft, ook de opgaven a) en b) goed heeft. Met andere woorden: men kan het wiskundig niveau van een respondent zien als een punt op een continuüm en men verwacht dat de respondent alle items goed heeft die aan de ene kant van dat punt liggen en alle items fout heeft die aan de andere kant van dat punt liggen. Wanneer men nu 'goed' en 'fout' vervangt door 'eens' en 'oneens', heeft men het patroon dat men verwacht bij een Guttmanschaal. De schaalwaarde is dan eenvoudigweg gelijk aan het aantal items dat men goed heeft of waarmee men het eens is.

Voorbeeld Guttmanschaal

Een onderzoek in de jaren '70 in de staat Connecticut. De attitude ten aanzien van abortus werd vastgesteld door een reeks omstandigheden voor te leggen die varieerden in de mate van extremiteit. Bij elke omstandigheid werd gevraagd of men voor of tegen abortus zou zijn. De resultaten staan hieronder.

Tabel 21: Verwerping van de omstandigheden waaronder abortus toelaatbaar is⁶

	Omstandigheid	Verwerping
1	carrière of opleiding wordt afgebroken	60%
2	te jonge leeftijd	57%
3	financiële problemen moeder	55%
4	te hoge leeftijd	55%
5	moeder wil het kind niet	54%
6	problemen om het kind alleen op te voeden	48%
7	zwangerschap/geboorte heeft negatieve psychologische effecten	32%
8	zwangerschap/geboorte heeft negatieve effecten voor gezondheid	28%
9	zwangerschap is het gevolg van verkrachting of incest	28%
10	genetische risico's voor het kind	25%
11	zwangerschap/geboorte is levensbedreigend voor de moeder	23%

Dus als een arts omstandigheid 6 verwerpt, verwachten we dat alle omstandigheden van 1 tot 6 ook verworpen worden. Als omstandigheid 7 acceptabel wordt geacht, verwachten we dat dit ook het geval is voor de omstandigheden 8 t/m 11.

Voorbeeld Guttmanschaal

Over het algemeen hebben verschillende rekenvaardigheden een opklimmende moeilijkheidsgraad.

Tabel 22: Opklimmende moeilijkheidsgraad van vaardigheden

OP	kinderen leren eerst 'optellen'
AF	en daarna 'aftrekken'
VE	vervolgens 'vermenigvuldigen'
DE	'delen'
MA	dan 'machtsverheffen'
WO	en vervolgens 'worteltrekken'

Het zal niet vaak voorkomen dat een kind eerst leert vermenigvuldigen en daarna optellen. Er is een duidelijke volgorde waarin de kinderen de rekenvaardigheden verwerven, en andersom kan men uit het feit dat iemand al kan delen, afleiden dat deze persoon ook zal kunnen optellen, aftrekken en vermenigvuldigen. In termen van Guttman: de rekenvaardigheden vormen een eendimensionale, cumulatieve schaal. Een dergelijke schaal geeft tegelijkertijd een ordening van items (stimuli) en van personen, in het voorbeeld: rekenvaardigheden en kinderen.

⁶ zie: Pligt, e.a. 1995, bl.186

tabel 23: Rekenvaardigheidsscores op een cumulatieve schaal⁷

	kinderen					
	F	B	E	A	D	C
OP	1	1	1	1	1	1
AF	0	1	1	1	1	1
VE	0	0	1	1	1	1
DE	0	0	0	1	1	1
MA	0	0	0	0	1	1
WO	0	0	0	0	0	1
	1	2	3	4	5	6

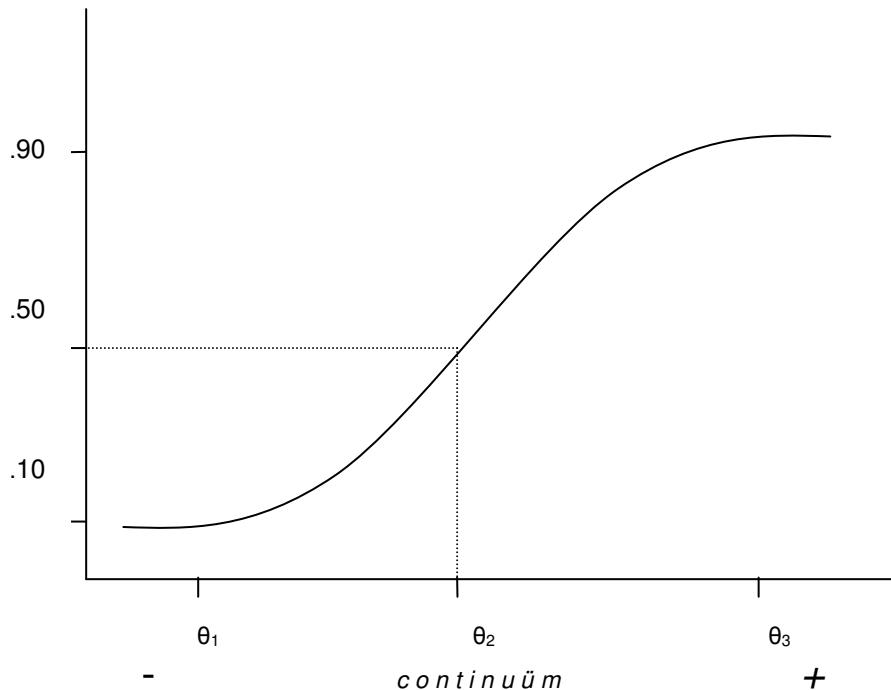
Een "1" betekend dat een kind de taak goed had, een "0" is fout. De rekenvaardigheden in deze tabel voldoen perfect aan de eisen van een cumulatieve schaal. We kunnen de rekenvaardigheden vervangen door attitude-uitspraken die elk een verschillende mate van gunstigheid uitdrukken. Als de antwoorden dan voldoen aan hetzelfde patroon als bij het rekenvoorbeeld, hebben we een cumulatieve schaal (die dan ook ééndimensionaal is: één onderliggende attitude uitdrukt).

Score op Guttmanschaal

Wanneer items een perfecte Guttman-schaal vormen kan in één score worden uitgedrukt met welke uitspraken de persoon het eens is en met welk oneens. Per respondent kan namelijk een keerpunt worden vastgesteld: met alle uitspraken boven een bepaalde gunstigheid is de persoon het niet meer eens en met de uitspraken onder het keerpunt is de persoon het wel eens. Elke respondent wordt namelijk verondersteld op het onderliggende attitude-continuum een positie in te nemen die we de schaalwaarde van de respondent noemen (θ : theta). In de nu volgende grafiek wordt dit geïllustreerd.

⁷ Ontleend aan: *Pligt e.a. p.186*

Figuur 5: Itemkarakteristiek net verschillende respondentposities



In bovenstaande grafiek heeft de respondent met schaalwaarde θ_1 10% kans om een positief antwoord te geven; die kans is voor de respondent met schaalwaarde θ_3 gelijk aan 90%. De respondent met schaalwaarde θ_2 heeft een kans van 50% om een positief antwoord op item i te geven. De schaalwaarde van een item is gedefinieerd als dat punt op het continuüm waarop 50% kans bestaat op een positief antwoord. In de grafiek is de schaalwaarde van de respondent met θ_2 gelijk aan de schaalwaarde van het item. Zo worden op hetzelfde continuüm zowel respondenten als items afgebeeld. Uit de aannamen van het Guttman model kan men afleiden dat wanneer alle items monotoon zijn op hetzelfde continuüm (dus eendimensionaal zijn) dan mogen de itemkarakteristieken elkaar niet snijden.

Per kruistabel één lege cel

Hiervoor is al gesteld dat als de items een Guttmanschaal vormen er per respondent kan een keerpunt kan worden vastgesteld en dan weten we welke vragen hij positief en welke hij negatief heeft beantwoord. Een implicatie hiervan is dat voor elk tweetal items geldt dat er in een kruistabel één lege cel is. Neem twee items i en j waarvan de schaalwaarden op eenzelfde attitude-continuüm liggen; de schaalwaarde van item i is hoger dan die van item j . Anders gezegd: item i is positiever dan item j . Wanneer we nu een kruistabel van deze twee items opstellen, verwachten we het volgende.

tabel 24: Kruistabel van item i en j

(item i is positiever dan item j; + is ja/eens en - is nee/oneens)

		item j	
		+	-
item i	+	leeg	gevuld
	-	gevuld	gevuld

Wanneer iemand 'nee' zegt op item j, kan hij niet vervolgens 'ja' zeggen op item i. Stel item i is: 'Ik vind het hier afwisselend' en item j is: Ik vind het hier heel eentonig'. Als iemand 'eens' op item j, kan hij vervolgens niet 'eens' zeggen op item i.

In de praktijk van het attitude-onderzoek blijkt het antwoordpatroon zelden zo mooi op te gaan: de 'lege' cel is gevuld, bevat dus afwijkingen. Het is namelijk moeilijk om een serie attitude-items te vinden die een monotone itemkarakteristiek hebben en die gespreid op het attitude-continuüm liggen. Men streeft naar items die de attitude in verschillende sterktes tot uitdrukking brengen. Wanneer men 'gematigde' items selecteert, is het heel wel mogelijk dat dergelijke items een non-monotone itemkarakteristiek vertonen. Dit versterkt de tendens om in Guttmanschalen dezelfde soort items op te nemen als in Likertschalen, namelijk extreem geformuleerde items.

Nadat we in het kader van een Guttmanschaalconstructie de data verzameld hebben, meestal dus antwoorden van respondenten op dichotome items, is het eerste punt dat zich voordoet het vinden van een rangorde van de items en van de respondenten. De rangorde van de items is de rangorde van het percentage positieve antwoorden binnen de steekproef en de rangorde van de respondenten is de rangorde van het aantal positief beantwoorde items. Stel we leggen 5 items voor aan 8 respondenten die per item aangeven of zij het wel of niet eens zijn met dat item. De resultaten zijn als volgt:

tabel 25: Scores van 8 respondenten op 5 dichotome items

pp	1	2	3	4	5	score
1	+	+	+	+	+	5
2	+	-	+	+	-	3
3	+	-	+	-	-	2
4	-	-	+	-	-	1
5	+	-	+	+	-	3
6	+	+	+	+	+	5
7	+	-	-	-	+	2
8	+	-	+	+	+	4
	7	2	7	5	4	25

We hergroeperen de tabel door de respondenten te rangordenen naar hun aantal positief beantwoorde items.

tabel 26: Gerangordende respondentenscores

pp	1	2	3	4	5	score
1	+	+	+	+	+	5
6	+	+	+	+	+	5
8	+	-	+	+	+	4
2	+	-	+	+	-	3
5	+	-	+	+	-	3
3	+	-	+	-	-	2
7	+	-	-	-	+	2
4	-	-	+	-	-	1
	7	2	7	5	4	25

Nu rangordenen we de items (kolommen) naar de volgorde van het aantal keren dat ze bevestigend werden beantwoord.

tabel 27: Gerangordende scores van items en respondenten

pp	1	3	4	5	2	score
1	+	+	+	+	+	5
6	+	+	+	+	+	5
8	+	+	+	+	-	4
2	+	+	+	-	-	3
5	+	+	+	-	-	3
3	+	+	-	-	-	2
7	+	-	-	+	-	2
4	-	+	-	-	-	1
	7	7	5	4	2	25

Uit voorgaande tabel kunnen we het volgende afleiden dat de volgorde van minst positieve naar meest positieve item is: item 2 (minst positief), item 5, item 4, item 3, item 1 (meest positieve item).

Hoe kunnen we nu vaststellen of aan de eisen van het model is voldaan? Want het perfecte response-patroon volgens Guttman (zie tabel 23) vinden we in de praktijk zelden. In dat perfecte response-patroon behoren we in elke 2 x 2 tabel een lege cel te vinden. In de praktijk zijn we toleranter en eisen dat het aantal in de 'lege;' cel klein moet zijn.

Om na te gaan of aan de eisen is voldaan, worden homogeniteitsindexen berekend voor:

1. elk tweetal items
2. elk item
3. de schaal

ad.1: Homogeniteitsindex voor elk tweetal items

Voor elk tweetal items wordt de homogeniteitsindex H_{ij} berekend. Elke H_{ij} dient groter dan nul te zijn.

<u>Homogeniteitsindex voor elk tweetal items</u>	
$H_{ij} = \frac{\phi}{\phi_{\max}} = 1 - \frac{e_{ij}}{e_{ij.kans}}$	
e_{ij}	aantal fouten in de 'lege' cel
$e_{ij.kans}$	aantal verwachte fouten in de 'lege' cel

Als de verwachte lege cel inderdaad geheel leeg is, dan is de H_{ij} gelijk aan 1. Als rekenvoorbeeld nemen we item 4 versus item 5. De kruistabel kunnen we afleiden uit tabel 27

tabel 28: Kruistabel van item 4 vs item 5

		item 5		
		eens	oneens-	
it 4	eens	3	2	5
	oneens	1	2	3
		4	4	8

Item 4 is positiever dan item 5, dus is het volgens het model niet mogelijk dat een respondent die 'nee' zegt tegen item 4 (het met dat item oneens is) vervolgens 'ja' zegt tegen item 5. Deze cel had dus leeg moeten zijn; echter in tabel 28 zien we dat deze cel één fout bevat, dus $e_{ij} = 1$. Het verwachte aantal fouten voor deze cel is: $(4 \cdot 3)/8 = 1,50$. Dus de H_{ij} is gelijk aan:

$$H_{45} = 1 - \frac{1}{\frac{4 \cdot 3}{8}} = 0,33$$

Op soortgelijke wijze kan men de overige kruistabellen afleiden en de H_{ij} 's berekenen. Deze staan in de nu volgende tabel.

tabel 29: H_{ij} 's van de kruistabellen die uit tabel 27 zijn af te leiden

		H_{ij}
1 vs 3:	$= 1 - 1/[(7 \cdot 1)/8]$	$= -0,14$
1 vs 4:	$= 1 - 0/[(5 \cdot 1)/8]$	$= 1$
1 vs 5:	$= 1 - 0/[(4 \cdot 1)/8]$	$= 1$
1 vs 2:	$= 1 - 0/[(2 \cdot 1)/8]$	$= 1$
3 vs 4:	$= 1 - 0/[(5 \cdot 1)/8]$	$= 1$
3 vs 5:	$= 1 - 1/[(4 \cdot 1)/8]$	$= -1$
3 vs 2:	$= 1 - 0/[(2 \cdot 1)/8]$	$= 1$
4 vs 5:	$= 1 - 1/[(4 \cdot 3)/8]$	$= 0,33$
4 vs 2:	$= 1 - 0/[(2 \cdot 3)/8]$	$= 1$
5 vs 2:	$= 1 - 0/[(2 \cdot 4)/8]$	$= 1$

In bovenstaande tabel is $H_{35} = -1$ en $H_{31} = -0,14$.

Item 3 zou moeten worden verwijderd, we geven de verder berekeningen met en zonder item 3 teneinde het effect van de verwijdering van een item te demonstreren.

ad. 2 Homogeniteitsindex voor elk item

Voor elk item wordt de H_i berekend (de H_i is te vergelijken met de item-rest-correlatie). Alle H_i 's dienen groter dan 0,30 te zijn.

$$H_i = 1 - \frac{\sum e_{ij}}{\sum e_{ij, \text{kans}}}$$

Voor item 5 verloopt de berekening als volgt. De kruistabellen met item 5 bevatten in totaal 2 fouten. Het verwachte aantal fouten van de kruistabellen van item 5 met

item 2 is $(2 \cdot 4)/8 = 1$
 item 4 is $(4 \cdot 3)/8 = 1,50$
 item 3 is $(4 \cdot 1)/8 = 0,50$
 item 1 is $(4 \cdot 1)/8 = 0,50$.

Het totaal aantal verwachte fouten voor item 5 is $= 3,50$ fouten. De H_i voor item 5 is nu:

$$H_5 = 1 - \frac{2}{3,50} = 0,43$$

Zo kan men ook de overige H_i 's berekenen.

Tabel 30: Hi's mét en zònder item 3

itemnr	H met item 3	H zonder item 3
1	0,56	1,00
2	1,00	1,00
3	0,11	--
4	0,71	0,65
5	0,43	0,67

ad. 3: Homogeniteitsindex voor de schaal

Tenslotte wordt voor de hele schaal de homogeniteitsindex berekend

$$H = 1 - \frac{\sum e}{\sum e_{kans}}$$

We berekenen nu de H over alle 5 items:

$$H = 1 - \frac{3}{\frac{8}{8} + \frac{6}{8} + \frac{2}{8} + \frac{2}{8} + \frac{12}{8} + \frac{4}{8} + \frac{4}{8} + \frac{5}{8} + \frac{5}{8} + \frac{7}{8}} = 0,56$$

Wanneer we item 3 weglaten, krijgen we voor de H:

$$H = 1 - \frac{3}{\frac{8}{8} + \frac{6}{8} + \frac{2}{8} + \frac{12}{8} + \frac{4}{8} + \frac{5}{8}} = 0,78$$

Door het weglaten van item 3 stijgt de H en wel van 0,56 tot 0,78.

Constructieprocedure van Guttmanschaal in stappen

1. Verzamel de data, eventueel antwoorden dichotomiseren.
2. Items en respondenten rangordenen van (zie tabel 16)
3. Bepaal de H_{ij} 's van de itemparen; indien H_{ij} negatief dan item verwijderen.
4. Bepaal de H_i 's en de H. Indien een of meer H_i 's kleiner dan 0,30 zijn dan item verwijderen. Alle H_i 's en de H moeten groter dan 0,30 zijn.

Swanborn refereert aan Mokken wanneer hij criteria (vuistregels) geeft voor de H_{ij} 's, de H_i 's en de H (Swanborn, 1988, p. 138):

"Alle H_{ij} 's dienen groter dan nul te zijn. Wanneer de H_i 's en de H groter dan 0,50 zijn, wordt gesproken van de goede schaal, indien tussen de 0,40 en 0,50 wordt gesproken van een matige schaal, tussen de 0,30 en 0,40 van een zwakke schaal en onder de 0,30 spreken we niet van een schaal."

Tot besluit van deze paragraaf over de Guttmanschaal sommen we in het kort enkele voor- en nadelen op van dit schaaltype:

- De schaalscore van de respondent bevat alle informatie over het antwoordpatroon op elk item mits toevalsfouten geen rol spelen. In de ontwikkelingspsychologie ken men de ordinale schaal van Uzgiris en Hunt. Deze vertoont eigenschappen die op een Guttmanschaal lijken: zie E.E.J. de Bruyn: 1993.
- De schaal is tevens een toets voor de eendimensionaliteit van de attitude. Dit voordeel wordt kleiner door de mogelijkheid dat toevalsfouten in de antwoorden het perfecte cumulatieve responsepatroon te niet doen.
- Een nadeel van de Guttmanschalingsmethode is de beperking tot eendimensionale attitude-domeinen.
Ten eerste: eendimensionaliteit is niet uitsluitend een eigenschap van een verzameling items, het is een patroon van een attitude voor een bepaalde populatie respondenten. Een bepaalde verzameling items kan in de ene populatie respondenten eendimensionaal zijn, maar hoeft dat niet te zijn in een andere populatie respondenten. Ook is het mogelijk dat een verzameling items op een bepaald tijdstip eendimensionaal is voor een bepaalde populatie personen, maar dat niet meer is enige tijd later.
Ten tweede: Het is moeilijk om attitude domeinen te vinden die eendimensionaal zijn. Neem bijvoorbeeld de houding van mensen ten aanzien van de overheid. De opinie over bepaalde juridische wetgeving kan heel verschillend zijn van de opinie over bepaalde belasting wetgeving.

9 Slotwoord schalen

In dit hoofdstuk zijn drie schaaltechnieken besproken. Zowel de Thurstoneschaal, de Likertschaal als de Guttmanschaal nemen aan dat attitude-items een bepaalde positie innemen op een (onderliggend) attitude-continuüm. Het verschil tussen deze drie schalen is als volgt:

- a. De Thurstoneschaal is gebaseerd op de aanname dat respondenten met een bepaalde plaats op het attitude-continuüm het alleen eens zijn met die items met een plaats op het continuüm die ongeveer gelijk is aan die van de respondent.
- b. De Likertschaal is gebaseerd op de aanname dat de kans om het met een item eens te zijn, geleidelijk toeneemt naarmate de plaats van de respondent op het attitude-continuüm die van het item nadert.
- c. De Guttmanschaal is gebaseerd op de aanname dat respondenten met een bepaalde plaats op het attitude-continuüm het eens zijn met de items die aan de ene kant van de plaats van de respondent 'liggen' en het oneens zijn met items die aan de andere kant van de plaats van de respondent liggen. Het verkregen meetniveau is een ordinale schaal.

Uit het feit dat bij gelijktijdige voorlegging aan respondenten van de Thurstoneschaal en de Likertschaal de correlaties met externe variabelen bij de Thurstone-schalen meestal wat lager zijn dan bij de Likertschalen, kan men afleiden dat de validiteit van de Thurstoneschalen in het algemeen wat lager is dan die van de Likertschalen. Wanneer echter het vinden van schaalwaarden van een aantal stimuli op zich het doel is, is de Thurstoneschaal de aangewezen schaaltechniek. Men kan hierbij denken aan bijvoorbeeld een beroepenprestigeschaal, en dergelijke.

Swanborn vergelijkt de Likertschaal en de Guttmanschaal en stelt:

"Laat ons vooropstellen dat er weinig of geen gevallen bekend zijn waarbij een bepaalde verzameling items wél als Likertschaal maar niet als Guttmanschaal, of omgekeerd, gebruikt kan worden. Bovendien als men van een bepaalde verzameling items zowel een Likertschaal als een Guttmanschaal maakt, correleren de beide scorereeksen over een verzameling respondenten meestal zeer hoog, veelal meer dan 0,90. Het praktisch nut van het gebruik van Guttmanschalen waar men ook Likertschalen kan toepassen -d.w.z. op het terrein van de attitude- en opiniëmeting- is dus vrij gering." (Swanborn, 1988, p. 202)

Swanborn meent dat de relevantie van de Guttmanschaal ligt bij het deterministische extreme geval en vooral op terreinen buiten de attitudemeting. Hij noemt als voorbeelden de Bogardusschaal voor sociale afstand, het stemgedrag van de rechters bij het Amerikaanse Hoogerechtshof. Voorts merkt Swanborn op prestatiedata zich uitstekend lenen voor het Guttmanmodel (zie Swanborn, 1988, bl. 205/206). Hierbij willen we opmerken dat ordinale schalen in de ontwikkelingspsychologie Guttmanachtige eigenschappen hebben (zie: De Bruyn, 1993).

Concluderend stellen we dat het veel werk is om een Guttmanschaal te construeren, terwijl de Likertschaal veel eenvoudiger is te construeren en de resultaten toch acceptabel zijn. Het is dan ook niet verwonderlijk dat met in sociaal-wetenschappelijk onderzoek veel vaker de Likertschaal aantreft.

Degenen die meer van schaaltechnieken willen weten wordt het boek van Swanborn (1987) aangeraden. Degenen die hun statistiekkennis willen ophalen, kunnen de twee boeken van Van Peet, Hox en Witten boer bestuderen (2004, 2005). Degenen die daarna echt zelf aan het onderzoekswerk willen gaan, wordt het uitstekende boek van Andy Field (2005) aangeraden. Dit boek is zeer toegankelijk geschreven en bespreekt voorts uitvoerig de SPSS-uitvoer.

Literatuur

Bruyn, E.E.J. de (1993). Psychodiagnostiek (Een systematische inleiding vanuit het klinische gezichtspunt). Rotterdam: Lemniscaat, 1985, 3e druk 1993.

Campbell, D.T. (1950). The indirect assessment of social attitudes. *Psychological Bulletin*, 47, p. 15-28.

Drenth, P.J.D & Sijtsma, K., (1990). Testtheorie. Houten/Antwerpen: Van Loghum.

Field, Andy (2005). *Discovering Statistics Using SPSS for Windows*. London, Sage Publications, 2005, second edition

Hoyle, R.H., M.J.Harris and C.M.Judd (2002). *Research Methods in Social Relations*. London: Wadsworth, Thomson Learning; Seventh Edition. ISBN: 0-15-506139-9

McIver, J.P. & E.G. Carmines (1981). *Unidimensional Scaling*. London, Sage Publications, nr. 24.

Peet, A.A.J.v. & G.L.H. van den Wittenboer & J.J.Hox (2004). *Toegepaste Statistiek, Beschrijvende Technieken*. Groningen: Wolters-Noordhoff/Senfert Kroese (1e druk 1995); tweede druk 2004.

Peet, A.A.J.v. & G.L.H. van den Wittenboer & J.J.Hox (2005). *Toegepaste Statistiek, Inductieve Technieken*. Groningen: Senfert Kroese, (1e druk 2001), 2e druk.

Pligt, J.v.d. & N.K.de Vries (1995). *Opinies en attitudes (metingen, modellen en theorie)*. Meppel: Boom.

Swanborn, P.G., (1987). *Methoden van Sociaal-Wetenschappelijk Onderzoek*. Meppel: Boom, 1971 (1981) nieuwe editie.

Swanborn, P.G., (1982). *Schaaltechnieken*. Meppel, Boom, 1982 (2e dr.1988).