

DIE VERDRAAIDE WERKELIJKHEID

Die verdraaide werkelijkheid

Hoe komt de onderzoeker aan zijn gegevens?

Over de natuur  
 Een filosoof (Kant, 1724-1804) studeerde psychologie in Am-  
 sterdam. Hij werkte achtereenvolgens bij de vakgroep Psychologie,  
 toevoeging van de Universiteit van Amsterdam, bij Psychologie aan de  
 Universiteit van Leiden en in Nederland van de Coelest psychologie-  
 en (een overzichtswerk van de psychologie). Hij werkte aan een  
 dissertatie over lichtheid, Over van Douchen (Duchon),  
 1841) heeft zich bezig met de psychologie van de mens en  
 tevens, was student-assistent medisch bij Psychologie in Le-  
 den en student-assistent medisch en natuurwetenschappen, bij Psy-  
 chologie in Leiden. Louis Thorens (1848) studeerde psychologie  
 aan de Universiteit van Amsterdam. In 1877 promoveerde hij daar  
 op een methodologische studie over het weten van onderwerpen  
 in de natuurwetenschappelijke onderzoek. Hij is als wetenschappelijk mede-  
 directeur verbonden aan de vakgroep Wiskunde en Letteren  
 Psychologie van de Universiteit van Leiden. Hij heeft zich vooral  
 bezig met onderzoek naar de ontwikkeling van ethische kennis  
 tussen ouders en jonges kinderen. Hij publiceerde over lichtheid,  
 in 1890, de methode van onderzoek naar de ontwikkeling  
 ontwikkeling in een wetenschappelijk onderzoek.

### Over de auteurs

*Leen Beukelaar* (Rotterdam, 1947) studeerde psychologie in Amsterdam. Hij werkte achtereenvolgens bij de vakgroep Psychofysiologie van de Universiteit van Amsterdam, bij Pedagogiek aan de Universiteit van Leiden en is nu redacteur van de *Codex psychologicus* (een overzichtswerk van de psychologie). Hij werkt aan een dissertatie over linkshandigheid; *Gert van Donselaar* (Den Haag, 1947) houdt zich binnen de psychologie bezig met 'methoden en technieken', was student-assistent statistiek bij Psychologie in Leiden en student-assistent statistiek en 'onderzoekspracticum' bij Pedagogiek te Leiden; *Louis Tavecchio* (1946) studeerde psychologie aan de Universiteit van Amsterdam. In 1977 promoveerde hij daar op een methodologische studie over het meten van onderwijsgedrag in de lichamelijke opvoeding. Hij is als wetenschappelijk hoofdmedewerker verbonden aan de vakgroep Wijsgerige en Empirische Pedagogiek van de Universiteit van Leiden. Hij houdt zich vooral bezig met onderzoek naar de ontwikkeling van affectieve relaties tussen ouders en (jonge) kinderen. Hij publiceerde over lichamelijke opvoeding, de meting van onderwijsgedrag en meetinstrumentontwikkeling in sociaal-wetenschappelijk onderzoek.

LEEN BEUKELAAR, GERT VAN DONSELAAR EN  
LOUIS TAVECCHIO

## Die verdraaide werkelijkheid

*Hoe komt de onderzoeker aan zijn gegevens?*

1. Inleiding	3
2. De vraagstelling	12
3. De veronderstellingen	24
4. De methode	78
5. Schets van de methode	88
6. De methode	111
7. De keuze van de methode	131
8. Kijkt men alleen naar alternatieve methoden?	144
Index	155
Register	177

1984-90 200 00 0000  
0000 00 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000

1984-90 200 00 0000  
0000 00 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000  
0000 0000 0000 0000 0000 0000

BASISBOEKEN/AMBO/BAARN

LEEN WERELDAR - GERT VAN DONSELAR EN  
LOUIS TAVOCHIO

## Die verdienste wetenschappelijk

Hoe komt de onderzoeker aan zijn gegevens?

ISBN 90 263 2042 6

copyright © 1980 by Uitgeverij Ambo bv, Baarn

ontwerp omslag: Guus Ros

typografie: Hans Gordijn

Basisboeken worden uitgegeven door

Uitgeverij Ambo bv, Baarn

Verspreiding voor België:

Uitgeverij Westland nv, Schoten

8.80.B38

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

No part of this book may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from the publisher.

## INHOUD

- Inleiding 9
1. Een onderzoek in vogelvlucht 19
  2. Observatie 25
  3. Het vraaggesprek 47
  4. De vragenlijst 64
  5. Inhoudsanalyse 79
  6. Schaalconstructie 96
  7. De test 113
  8. De keuze van de methode 131
  9. Kijken naar feiten: een alternatieve onderzoeksmethode 144
- Besluit 155
- Register 157

## VOORWOORD

Geen enkel boek komt tot stand zonder de hulp, inspiratie en kritiek van anderen. Allereerst gaat onze dank uit naar de leden van de vakgroep Wijsgerige en Empirische Pedagogiek van de Rijksuniversiteit van Leiden. Vooral bij het ontstaan van de syllabus, die de basis van dit boek vormt, maar ook later zijn zij ons voortdurend tot steun geweest.

Daarnaast bedanken wij Letty Beukelaar-Santé, Marijke Cassee, Willem de Ruyter, Harry Vorst en Pieter Vroon. Veel meer fouten en veel minder waardevols zou u gevonden hebben, hadden zij ons niet geholpen.

Amsterdam, februari 1980

## INLEIDING

De wereld wordt overstroomd met resultaten van allerlei soorten onderzoek. Deze resultaten worden gepresenteerd en veelal geaccepteerd als feiten. Vooral degenen die een opleiding volgen, krijgen daar mee te maken, maar ook 'gewone' krantelers en televisie-kijkers worden geconfronteerd met wetenschappelijke resultaten. Zelden wordt door hen aan de resultaten getwijfeld. Veel betogen wordt kracht bijgezet door het vermelden van percentages; veel discussies worden gewonnen door degene die de meeste onderzoeksresultaten aan kan halen. Als iemand een bepaalde mening verkondigt en ten antwoord krijgt: 'Maar recent onderzoek heeft aangetoond dat 83% van de Nederlanders het niet met u eens is!' of 'Maar onderzoek aan de universiteit van Minnesota heeft toch aangetoond dat dat niet waar is!' zwijgt de eerste spreker meestal bedremmeld. Dat is in veel gevallen niet terecht. Onderzoeksresultaten, of ze nu gepresenteerd worden door wetenschappelijke of door commerciële organisaties, zijn vaak in de verste verte geen feiten. Een voorbeeld kan dat duidelijk maken.

Zowel de Consumentenbond als de Bovag hebben onafhankelijk van elkaar een onderzoek gedaan naar de tevredenheid van klanten van garages. Volgens het onderzoek van de Consumentenbond was de grote meerderheid van de klanten ontevreden, de Bovag daarentegen vond juist dat het omgekeerde het geval was (beide percentages lagen rond 80%; volgens de Consumentenbond was dus 80% ontevreden, volgens de Bovag was 80% tevreden). Ogenscheinlijk een onoplosbaar probleem. De vertegenwoordigers van beide organisaties kwamen er tijdens een televisie-uitzending dan ook niet uit. Toch is de oplossing niet zo moeilijk.

Eén van de dingen die resultaten sterk kunnen beïnvloeden is de manier waarop deze tot stand komen. In dit geval bleek de Consumentenbond iemand ontevreden te noemen als hij in de afgelopen drie jaar éénmaal ontevreden was geweest, de Bovag daarentegen noemde iemand ontevreden als hij over het algemeen ontevreden was over de behandeling. Nu zijn de gevonden resultaten ineens niet meer met elkaar in tegenspraak. Als we aannemen dat er verder geen dingen zijn gebeurd die de uitslag vertekenen (bij nadere

analyse zou ongetwijfeld blijken dat die aanname ten onrechte is gedaan), komen we tot de volgende conclusie: 20% van de garagebezoekers had over de behandeling gedurende de laatste drie jaar geen enkele klacht, een andere 20% was vrijwel voortdurend ontevreden en 60% was over het algemeen wel tevreden, maar had minstens één klacht over de behandeling in een garage gedurende de laatste drie jaar.

Alleen doordat hier twee onderzoeken over hetzelfde onderwerp met elkaar konden worden geconfronteerd, was het mogelijk te ontdekken dat de resultaten zoals die gepresenteerd werden niet juist waren. Meestal is er echter maar één onderzoek of zijn er meerdere, die elkaar niet tegenspreken. De enige tegenspraak die er dan kan komen, moet van mensen of groepen afkomstig zijn die het predikaat 'wetenschappelijk' niet willen of mogen voeren. We kunnen daarbij denken aan bewonerscomités en allerlei actiegroepen. Als deze 'leken' zich tegen de wetenschap uitspreken, zien we dat door grote groepen, politici vaak voorop, blindelings de kant van de wetenschap wordt gekozen. (We zien overigens dat de laatste tijd die actiegroepen zich versterken met doctorandussen en ingenieurs, om een grotere kans te maken (en met succes) om gehoord te worden. In hoeverre ze zich daardoor overgeleverd hebben aan de 'wetenschappers' om de wetenschap te bestrijden, is nog een open vraag.)

Dat de wetenschap die 'blindelingse' voorsprong heeft, zien we heel duidelijk bij het onderwijs op openbare lagere scholen. Zoals iedereen weet, zijn er twee gangbare theorieën over het ontstaan van de verschillende plante- en diersoorten en de mens: de schepping en de evolutie. De tweede is de wetenschappelijke. In beide theorieën moet je echter evenzeer geloven, want ook de evolutie is niet bewezen. In 1977 is een boek verschenen waarin de belangrijkste ontwikkelingen staan binnen de erfelijkheidsleer. De conclusie van de schrijver is dat we op het moment zeer veel weten en kunnen verklaren over de verschillen binnen de soorten (bijvoorbeeld alle verschillende hondesoorten), maar in feite nog niets weten over de mechanismen die het mogelijk zouden maken dat de ene soort uit de andere ontstaat. Zo wordt er al honderden, zo geen duizenden jaren, gefokt met honden (een versnelde evolutie). Zeer veel en sterk uiteenlopende honderrassen zijn het resultaat. Toch kunnen alle honderrassen onderling gekruist worden, zij het soms met aanzienlijke fysieke moeilijkheden. Ze behoren dus tot dezelfde soort. Noch in theorie noch in de praktijk zijn er bewijzen voor



de juistheid van de evolutietheorie. Toch worden evolutie en schepping binnen het onderwijs zeer verschillend behandeld. De evolutietheorie is een verplicht vak, terwijl de scheppingstheorie alleen verkondigd wordt aan die kinderen wier ouders daar geen bezwaar tegen hebben. Het omgekeerde, dat toch evenzeer verdedigd zou kunnen worden, is onmogelijk.

De vraag die we in deze inleiding zullen proberen te beantwoorden is: Hoe komt het dat een uitspraak eerder aanvaard wordt als er 'wetenschappelijk' voor staat. En vervolgens: 'Is dat terecht?'

Om deze vragen te kunnen beantwoorden moeten we eerst weten wat dat 'wetenschappelijk' eigenlijk betekent. Hoewel in de verschillende boeken die hierover geschreven zijn, zeer veel en vaak verschillende eisen staan waaraan iets moet voldoen om wetenschappelijk te zijn, komen de volgende twee eisen vrijwel steeds voor:

1. Wetenschap zoekt naar de waarheid.
2. Wetenschap dient de vooruitgang van de mensheid.

We zullen in het onderstaande aantonen dat dit slechts reclameleuzen zijn, die net als alle andere reclame uitstekend werken ter stimulering van de verkoop, maar als omschrijving van het te verkopen produkt niet erg voldoen.

### *De waarheid*

De belangrijkste en door vrijwel alle wetenschappers aangehangen regel is: 'Een uitspraak is waar, als hij geldt voor een aantal gevallen waarvoor hij bedoeld is en er (nog) geen gevallen bekend zijn waarvoor hij ook zou moeten gelden maar waarvoor hij niet opgaat.' Een uitspraak die waar is, is bijvoorbeeld: 'Mensen worden niet ouder dan 200 jaar.' Er zijn immers geen (goed gedocumenteerde) gevallen bekend die niet aan deze uitspraak voldoen. Een bekende pedagogische uitspraak dat het mensenkind bij zijn geboorte het minst van alle zoogdieren op zijn volwassen vorm lijkt, is daarentegen niet waar. Hoewel er vermoedelijk meerdere voorbeelden zijn, lijkt in ieder geval de baby van een reuzenkangoeroe veel minder op de volwassen vorm dan een mensenbaby. De volwassen reuzenkangoeroe is bij benadering even groot als een volwassen mens. Mensenbaby's zijn bij hun geboorte echter bijna 20 maal zo groot als de kangoeroebaby (die is nl. slechts 2,5 cm lang). Ook de verhouding tussen hoofd en lichaam verschilt bij de kangoeroe meer dan bij de mens.

Op het eerste gezicht lijkt deze regel heel goed. In de praktijk blijkt dat echter tegen te vallen. In de eerste plaats blijken er vrijwel geen (of misschien wel helemaal geen) uitspraken te zijn die niet door een of ander onderzoek worden tegengesproken. Dit betekent dat er in feite helemaal niets waar is. In de tweede plaats, gedeeltelijk om de gevolgen van het eerste punt te verbergen, wordt deze regel hoogst zelden toegepast. Om te voorkomen dat men een uitspraak of theorie moet weggooiden, past men een aantal trucs toe. Men kan zijn uitspraak bijvoorbeeld een beetje veranderen. Bovengenoemde pedagoog kan aan zijn bewering toevoegen: behalve buideldieren. Slechts door nog veel meer gevallen te ontdekken, zouden we hem misschien van het idee af kunnen brengen dat de mensenbaby iets bijzonders heeft, in tegenstelling tot die van de dieren.

Er zijn echter nog meer manieren om van vervelende uitspraken af te komen die niet met de jouwe stroken. Men kan ze onder andere negeren. We zien dat bijvoorbeeld binnen het intelligentie-onderzoek. Onder de onderzoekers naar intelligentie kan men twee groepen onderscheiden. De ene is ervan overtuigd dat de verschillen in intelligentie tussen de mensen voor minstens 80% worden bepaald door erfelijkheid, de andere dat die verschillen hoofdzakelijk bepaald worden door de omgeving (opvoeding, school en dergelijke). De leden van de eerste groep gebruiken vaak resultaten van onderzoek naar lengte als vergelijkingsmateriaal voor hun intelligentie-onderzoek. Ze beweren dat lengte erfelijk bepaald wordt en dat, als ze dezelfde uitkomsten vinden met betrekking tot intelligentie, deze dan ook erfelijk bepaald moet zijn. De tegenwerpingen van hun tegenstanders worden hierbij volledig genegeerd. Zo zijn genetici er helemaal niet van overtuigd dat lengte erfelijk bepaald wordt mede door het verschijnsel dat kinderen gemiddeld langer zijn dan hun ouders. (Binnen de genetica mogen er wel verschillen zijn tussen ouders en kinderen, maar dat het gemiddelde verandert kan men genetisch niet verklaren.)

Nog kwalijker wordt het als de onderzoeker zijn uitspraak verdedigt door niet de uitspraak van zijn tegenstander aan te vallen, maar die tegenstander zelf. Zo worden sommige artikelen van tafel geveegd met het argument dat de schrijver ervan te stom is om zich in de discussie te mengen.

We zien dus dat de waarheidsregel, zoals we die hierboven hebben weergegeven, niet altijd wordt toegepast. Toch hebben wetenschappers, elk individueel, een mening over wat wel en wat niet waar is. Je kunt je afvragen of ze dan misschien over andere regels beschik-

ken waar ze gebruik van maken. Er zijn wel andere regels geweest. De voorloper van bovengenoemde regel luidde: 'Een uitspraak is waar als hij geldt voor alle gevallen waarop hij betrekking heeft.' Deze regel is zeer moeilijk te controleren, omdat men meestal uitspraken doet over zeer vele gevallen. Het is dan niet mogelijk die allemaal te controleren.

Een nog andere regel is dat iets wel waar zal zijn als iedereen het daarover eens is. Helaas is er voor zover wij weten niets waar iedereen het over eens is (uitgezonderd triviale uitspraken zoals vissen leven in het water).

Tenslotte is er nog een regel die eigenlijk uit het dagelijks leven komt, namelijk: 'Iets is waar als een expert (dat is iemand die het kan weten) het zegt.' Hoewel deze regel door de wetenschappers wordt verfoeid, is dit toch de regel die het meest gehanteerd wordt.

Ter illustratie het volgende voorbeeld:

Een van de schrijvers had andere resultaten gevonden dan een bekende onderzoekster op hetzelfde gebied. Toen hij haar daarover aansprak op een congres, vroeg zij hoelang hij zich al met dit onderwerp bezighield. Dat was ongeveer 2 jaar. Haar conclusie was toen dat zijn resultaten dan niet goed konden zijn, want zij was al 15 jaar met dit onderwerp bezig.

Maar hoe word je een expert en wat is dat eigenlijk. Een expert is iemand die een aantal anderen om zich heen verzameld heeft die erkennen dat hij een expert is.

Men kan dat op verschillende manieren bereiken. De meest gehanteerde manier is erkenning zoeken van een reeds erkende expert. Gedienstig zijn en niet tegenspreken op belangrijke punten zijn een paar van de ook onder ambtenaren bekende manieren om deze status te bereiken. Een andere manier is het om je heen verzamelen van mensen die min of meer van je afhankelijk zijn. Het zich bekwamen in het schrijven van subsidie-aanvragen kan hierbij zeer goed helpen. Een ook veel gebruikte manier is het opzetten van een grote mond. De hardste schreeuwer wordt vaak het meeste beluisterd. Als iemand maar hard genoeg verkondigt dat hij het beste onderzoek doet, gaat hij het niet alleen zelf geloven, maar overtuigt hij anderen ook.

We zien dat binnen de wetenschap, net als daarbuiten trouwens, het waarheidsgehalte van een uitspraak wordt afgemeten aan de status van de spreker.

Een bijkomend verschijnsel is dat, als men op een bepaald gebied erkend wordt als expert, er vaak gedacht wordt dat men op andere

gebieden óók expert is. Zo goed als acteurs alles over auto's en wasmiddelen schijnen te weten, zangeressen experts zijn op het gebied van zachte borrels, en voetballers en bejaarde D.J.'s ook nog verstandige rokers zijn, blijken uitspraken van gevestigde wetenschappers op gebieden die niets met hun vak te maken hebben ook als waardevol beschouwd te worden: onderwerpen als astrologie en parapsychologie worden vaak als nonsens afgedaan door wetenschappers die op dat gebied in het geheel niet door kennis worden gehinderd.

### *De vooruitgang*

De tweede eis die aan de wetenschap gesteld kan worden, zo hebben we hierboven gezien, is die voor het bevorderen van de vooruitgang. Als we op deze kwestie willen ingaan, moeten we een onderscheid maken tussen de vooruitgang van de wetenschap zelf en de vooruitgang van de mensheid.

Vooruitgang houdt in dat de situatie beter zal worden dan zij is, dat er een bepaald doel is; immers een doel nastreven sluit achteruitgang, zijwaarts gaan en voortgaan (bewegen zonder doel) uit. De vooruitgang van de wetenschap zouden we af kunnen meten aan het aantal problemen dat opgelost wordt.

Als we de vooruitgang van de mensheid er bij willen betrekken, wordt het echter een stuk moeilijker. Dit soort vooruitgang wordt gedefinieerd door de politiek, niet door de wetenschap. De politici zijn het echter in het geheel niet met elkaar eens over wat vooruitgang is. Wat voor de één een stap in de goede richting is, is voor de ander een heilloze weg.

De wetenschapper kan twee dingen doen. Hij kan in de eerste plaats zijn eigen politieke opvattingen trouw blijven. We zien dit bij de reeds aangehaalde onderzoekers op het gebied van intelligentie en erfelijkheid. Er schijnt een tamelijk sterk verband te bestaan tussen het milieu waaruit de onderzoeker afkomstig is en de theorie die hij aanhangt. Mensen uit de hogere milieus hangen vooral de erfelijkheidstheorie aan, terwijl onderzoekers uit de lagere milieus meer geporteerd zijn voor de omgevingsopvatting. Op het moment dat Nederland geregeerd wordt door een centrum-rechtse regering zien we dan ook onderzoekers die tot de eerste groep behoren, optreden als adviseurs, met als gevolg dat veel schoolexperimenten worden afgeremd of in de ijskast gezet (het heeft immers toch geen zin domme milieus betere scholen te geven), terwijl tijdens een cen-

trum-linkse coalitie de aanhangers van de omgevingstheorie de kans krijgen: het gevolg is dat ettelijke nieuwe schooltypen worden ingevoerd.

De tweede opvatting is de zo-de-wind-waait-waait-mijn-jasje-filosofie. Van deze keuze worden psychoanalytici nog wel eens beschuldigd. Men zou zich in dit geval kunnen voorstellen dat iemand tussen 1933 en 1945 in Duitsland onder behandeling is wegens het ontbreken van antisemitische gevoelens, terwijl hij, indien genezen, vanaf 1945 op de sofa ligt omdat hij die gevoelens juist wel heeft.

Een derde mogelijkheid is te zeggen dat wetenschap en politiek niets met elkaar te maken hebben, immers penicilline werkt even goed bij PvdA-stemmers als bij die van de VVD. Dat is natuurlijk waar, maar dat geldt even zeer voor zyclon-B en de neutronenbom. De opvatting dat de gebruiker verantwoordelijk is voor de gevolgen en niet de uitvinder zou wel eens onjuist kunnen zijn, zie bijvoorbeeld de schuldgevoelens van o.a. Einstein met betrekking tot de atoombom. Zelfs al heeft men een wat ouder kind honderd maal gewaarschuwd niet van papa's pilletjes te snoepen, dan toch is men verantwoordelijk als het kind het wèl doet, omdat men de pilletjes binnen zijn bereik heeft laten liggen. Elke wetenschapper stelt een politieke daad bij het openbaar maken van zijn bevindingen en blijft daar, in ieder geval deels, verantwoordelijk voor. Leonardo da Vinci beseftte dit terdege en openbaarde dan ook een aantal oorlogswerktuigen, die hij had uitgevonden, niet.

Lang niet alle produkten zijn echter duidelijk positief of negatief te noemen. Was dat wel zo geweest, dan hadden we met een aantal eenvoudige maatregelen de zaak nog ten goede kunnen keren. Een produkt als de auto heeft minstens zoveel voors als tegens. Natuurlijk maken we een hoop tijdwinst. Erasmus moest dagenlang zijn tonnetje met boeken voortschoppend lopen om een boek in de bibliotheek van Keulen te raadplegen. Nu kunnen we daar in enkele uren komen. Daarentegen schakelde Erasmus zijn verstand niet uit op het moment dat hij zich in beweging zette, zoals we veelvuldig tegenkomen bij de hedendaagse automobilist. Ook is het prachtig dat we nu overal naar de natuur kunnen gaan kijken, waarbij we dan maar vergeten dat we wel weg moeten, omdat in de omgeving waar we wonen de natuur begraven ligt onder grijze, langgerekte asfaltgrafstenen. Het is zelfs al zo erg dat delen van de te bezoeken pittoreske gebieden moeten verdwijnen om ze voor de automobilisten toegankelijk te maken (bijvoorbeeld de Mergelland-route in

Limburg). Hoewel er blijkens de reclame zeer veel mensen blij zijn dat ze rijden, wordt deze blijheid vermoedelijk in doden maar zeker in zwaargewonden duurder betaald dan de nu zozeer verafschuwde blijheid van de Romeinen in de amfiteaters. Hoewel de afschaffing van de auto op dit moment zeker tot nog veel meer narigheid aanleiding zou geven, toont het voorgaande wel aan dat de welzijnsstaat zeker niet dichterbij is gekomen.

Uit het bovenstaande kunnen we slechts één conclusie trekken, namelijk dat de wetenschap als geheel geen eensluidend antwoord kan geven op de vraag waar we naartoe moeten als we vooruitgaan en er zodoende dus ook niet aan kan meewerken.

Ook de vraag of de wetenschap vooruitgaat, dat wil zeggen raad weet met steeds meer problemen, is moeilijk te beantwoorden. Uiteraard is het zo dat we steeds andere vragen kunnen beantwoorden en dat we telkens andere antwoorden op bestaande vragen vinden, die ons beter bevallen, maar het is ook zo dat een aantal vragen waar vroeger wel een bevredigend antwoord op was, niet meer gesteld worden.

We kunnen dit illustreren aan een reclameboodschap van de Shell. Daarin wordt een verbeterde motorolie aangeprezen. Een kritische acteur heeft Shell, naar zijn zeggen, gevraagd waarom die olie verbeterd moest worden en of die oude dan niet goed meer was. Shell schijnt geantwoord te hebben dat die andere olie best was, maar dat nieuwe auto's hogere eisen stellen dan de oude en dat daarom die nieuwe olie nodig was. Dit antwoord bleek bevredigend, want van een verdere discussie wordt geen verslag gedaan. Minstens één interessante vraag ontbreekt, namelijk: is die nieuwe olie wel goed voor oude auto's. Niets in het verhaal wijst daarop, immers de oude was prima.

Iets dergelijks zou kunnen gelden voor de wetenschap. We weten niet méér, maar we weten iets anders. Hoewel het eerste mannetje op de maan het nog over een reuzenstap voor de mensheid had, is de mening van Feyeraabend hierover iets anders: 'Er waren miljarden dollars, duizenden goed getrainde assistenten, jaren van hard werk nodig om ons in staat te stellen enige onduidelijk sprekende en tamelijk beperkte tijdgenoten enkele weinig sierlijke huppeltjes te laten maken op een plaats waar niemand, die goed bij zijn verstand is, aan zou denken om naartoe te gaan: een uitgedroogde, luchtloze, hete steen. Terwijl mystici, slechts gebruik makend van hun geest, door hemelse sferen reisden tot God zelf, die zij aanschouwden in al zijn glorie, kracht ontvangend om hun leven voort te zetten en

verlichting ontvangend voor zichzelf en hun medemensen.'

Op dit moment wordt minachtend teruggekeken op de kennis die men honderd jaar geleden bezat. Hetzelfde deden echter de mensen van honderd jaar geleden. Er is dan ook geen enkele reden om aan te nemen dat de mensen over honderd jaar anders over ons zullen denken. De hele technologie van computers, ruimtevaart enzovoorts, is dan misschien even interessant voor hen als de resultaten van het zoeken naar de steen der wijzen in de middeleeuwen nu voor ons.

Men zou misschien dan ook beter over de voortgang van de wetenschap kunnen spreken dan over de vooruitgang. (Misschien zou men zelfs kunnen zeggen dat wetenschap geen doel heeft, maar een doel is. Zie ook hfdst. 9.)

### *Het beheersen*

Een facet dat in dit verband nog aangestipt dient te worden, is de volgorde waarin de kennis zich ontwikkelt. Veelal wordt daarvoor aangegeven: verbazing – begrijpen – voorspellen – beheersen. Dit gaat als volgt in zijn werk: Men verbaast zich over het feit dat Jantje nog steeds in zijn bed plast. Plotseling begrijpt men dat dat komt omdat men Jantje niet laat plassen voordat hij naar bed gaat. Men voorspelt dan dat het niet meer voor zal komen, als men hem wel laat plassen. Komt die voorspelling uit, dan kan men dit gedrag van Jantje voortaan beheersen. Men kan hem al naar men belieft al of niet in bed laten plassen.

Aan dit beheersen, als doel van de wetenschap, zitten echter enige belachelijke en gevaarlijke kanten. Belachelijk is dat we onderzoek doen naar dingen die we nooit zullen beheersen. We begrijpen eb en vloed, we kunnen ze perfect voorspellen, maar beheersen is er echt niet bij. We weten nu dat de aarde om de zon draait en begrijpen ook waarom, toch zijn er naar ons weten nog steeds geen plannen om de omgekeerde veel rechtvaardiger situatie te bewerkstelligen. Veel enger wordt het als we deze regel blindelings zouden (laten) opvolgen ten aanzien van dingen die we wel kunnen beheersen. Mochten onverhoopt de onderzoekers die verklaren dat intelligentie vrijwel geheel erfelijk bepaald is toch gelijk krijgen, dan zou dit leiden tot het verstrekken van fok-vergunningen op vertoon van een door een psycholoog getekende verklaring dat de intelligentie van die persoon toereikend is. Dat dergelijke praktijken tot de mogelijkheden behoren heeft een recent verleden ons geleerd.

## Conclusie

Uit het bovenstaande zou men kunnen afleiden dat alle wetenschappers boosaardige en geld- en statusbeluste figuren zijn, dat ze elke leugen aangrijpen die hen dichter bij hun doel brengt. Dat is uiteraard niet waar. Aan de andere kant is het beeld van een zichzelf wegcijferende wetenschapper die slechts het goede voor het volk voor ogen heeft evenmin juist. Het gedrag van de wetenschapper wordt door beide motieven beïnvloed. Voor de één weegt het eigen belang wat zwaarder, voor de ander het algemeen belang.

Het gevolg is wel dat de afgeleverde producten, de wetenschappelijke resultaten, niet zonder meer geaccepteerd kunnen worden. Men koopt tenslotte ook niet zo maar een stoel, omdat de meubelmaker zegt dat hij lekker zit. Stoelen worden uitgebreid bezeten, voordat men ze bezitten wil. Een probleem hierbij is dat wel iedereen weet hoe een stoel moet zitten om comfortabel te zijn, maar dat het beoordelen van onderzoek volgens velen een academische opleiding vereist.

Dat dat niet noodzakelijk is hopen we met dit boekje te bewijzen. We zullen in het bijzonder de fase van het verzamelen van gegevens (de dataverzameling) behandelen.

## Aanbevolen literatuur

Delen van de hieronder genoemde literatuur zullen u bekend voorkomen.

P. Feyerabend, *Science in a free society*. London, NLB, 1978.

H. Koningsveld, *Het verschijnsel wetenschap, een inleiding tot de wetenschapsfilosofie*. Meppel, Boom, 1976.

P. A. Vroon, *Intelligentie*. Baarn, Ambo, Basisboeken, 1979.

A. D. de Groot, *Methodologie*. Den Haag, Mouton, 1972.

P. A. Vroon, *Bewustzijn, hersenen en gedrag*. Baarn, Ambo, Basisboeken, 1976, 1979<sup>2</sup>.



## 1. EEN ONDERZOEK IN VOGELVLUCHT

Alle wetenschap begint met verbazing, zo ook alle onderzoek. Over deze verbazing zijn geen regels te geven. Over de stap om die verbazing om te zetten in een onderzoek dient een onderzoeker daarentegen wel degelijk verantwoording af te leggen. Die beslissing is vaak politiek getint. Men kan onderzoek gaan doen waar alleen een kleine, bevoorrechte groep iets aan heeft, of men kan iets doen waar brede lagen van de bevolking profijt van hebben. Vooral de verschaffer van het geld, meestentijds is dat de overheid, de hele bevolking dus, zou in deze fase iets in de pap te brokkelen moeten hebben. Dat komt er vrijwel nooit van, deels omdat de bevolking, evenals haar vertegenwoordigers, zich er ten onrechte niet voor interesseren, deels omdat de onderzoeksvoorstellen in een soort geheimtaal geschreven worden, en deels omdat de onderzoeker zich niet wenst te laten beïnvloeden. Dit laatste komt er op neer dat de bevolking gedurende lange tijd zeer grote bedragen moet geven aan mensen die zich niet verantwoorden en vaak dingen doen en uitzoeken om te bewijzen dat die geldschieter ongelijk heeft. Als men daarnaast nog nagaat dat een groot deel van het onderzoek nooit gepubliceerd wordt en dat nog eens de helft van het wel gepubliceerde onderzoek nooit meer door iemand aangehaald wordt, kan men zich niet aan de indruk onttrekken dat al dat geld op een betere manier besteed had kunnen worden. Om dit te verbeteren zouden in de eerste plaats de wetenschappers moeten leren hun ideeën in gewone mensentaal te verwoorden en in de tweede plaats zouden er systemen moeten komen tot openbaarheid met de mogelijkheid kritiek uit te oefenen. In de derde plaats zullen *alle* belanghebbenden zich meer voor onderzoek moeten gaan interesseren.

Een andere manier is het sterk uitbreiden van de wetenschapswinkels. Dit houdt in dat de bevolking zich verbaast en dat de wetenschappers over dat onderwerp, na overleg, een onderzoek doen.

### *Het begrijpen*

Nadat men op een al of niet verantwoorde manier tot de beslissing is gekomen over de vraag welk probleem men wenst te onderzoeken,

moet men erachter zien te komen hoe dat probleem in elkaar zit. Hiervoor zijn een aantal manieren:

a. Men heeft meestal zelf wel een bepaalde opvatting. Hoewel men zich hierop niet moet blindstaren, kan het een eerste begin vormen.

b. Er is vrijwel geen onderwerp waarover nog nooit iemand iets heeft geschreven. Hoewel er over veel onderwerpen zoveel is geschreven dat het bijna een mensenleven zou duren dat allemaal te lezen (over schizofrenie zijn bijvoorbeeld meer dan 10.000 boeken en artikelen verschenen), zijn er vaak overzichtswerken samengesteld en dient men zich in ieder geval van de meest recente ontwikkelingen op de hoogte te stellen (bijvoorbeeld in openbare of universiteitsbibliotheken).

c. Vaak zijn er mensen die zonder dat ooit op schrift te hebben gesteld, deskundigen zijn op het te onderzoeken gebied. Men kan hierbij denken aan leraren, artsen, sociaal werkers maar ook aan minder geleerde dames en heren. Om bijvoorbeeld iets over de sociale structuur van een wijk te weten te komen, kan men vaak te rade gaan bij de winkeliers in zo'n wijk, over gedrag in openbare plaatsen kan men zeer veel leren van marktkooplieden.

d. Men kan een verkennend onderzoek doen. Dit houdt in dat men, indien de bovengenoemde methoden niet mogelijk, gewenst of bevredigend zijn, zonder duidelijke ideeën naar de mensen of situaties gaat waarin het probleem zich voordoet.

De resultaten van deze fase zijn één of meer duidelijk geformuleerde verwachtingen over hoe de vork in de steel zit.

Laten we als voorbeeld iemand nemen die zich erover verbaast dat er vooral de laatste tijd zoveel geklaagd wordt over de behandeling in ziekenhuizen. Omdat hij zelf nooit in een ziekenhuis heeft gelegen, heeft hij geen idee hoe dat zou kunnen komen. Ook in de literatuur kan hij niets vinden dat hem op het juiste spoor kan zetten. (Dit kan twee redenen hebben: een onwaarschijnlijke, namelijk dat er echt niets over is geschreven, en een veel waarschijnlijker, dat hij niet goed heeft gezocht.) Vervolgens gaat hij er met een aantal 'deskundigen' over praten: een oom en een nicht die in het ziekenhuis hebben gelegen, twee vriendinnen van zijn zoon die beiden verpleegster zijn en een zwager van zijn vrouw die als internist in een ziekenhuis werkt. Dit maakt het beeld meteen veel duidelijker. Volgens de patiënten hebben de verple(e)g(st)ers en de artsen het op hun geweten, volgens de verple(e)g(st)ers de artsen en de patiënten en volgens de arts de patiënten en de verple(e)g(st)ers.

In zijn onwetendheid had zelfs hij deze mogelijkheden overwogen. In de verwachting dat andere deskundigen ongeveer dezelfde mening zouden hebben, besluit hij zelf maar eens een kijkje te gaan nemen. Achtereenvolgens probeert hij tevergeefs als arts, verpleger en schoonmaker in dienst van het ziekenhuis te komen. Als hem vervolgens ook de toegang tot het ziekenhuis, buiten de bezoeken, wordt ontzegd, treft hem bij het verlaten van het hospitaal het (on)geluk een been te breken, zodat hij alsnog geruime tijd in het ziekenhuis mag vertoeven. Door gedurende zijn opname goed om zich heen te kijken, en met de andere patiënten en, voor zover mogelijk, met het verplegend personeel en de artsen te spreken krijgt hij langzamerhand een idee waarom de sfeer binnen de ziekenhuizen de laatste tijd zo verslechtert. Door de vrij lage beloning die het verpleegkundig personeel ontvangt is er een groot verloop. Velen gaan naar particuliere instellingen die beter betalen. Hierdoor moeten meer leerlingen aangetrokken worden, waardoor de kwaliteit van de verzorging afneemt. Vandaar dus de ontevredenheid.

### *Voorspellen*

Onze onderzoeker heeft nu een idee, enig begrip van het probleem. Deze 'theorie' kan echter zeer persoonlijk zijn of slechts geldig in het ziekenhuis waar hij gelegen heeft. Alom in den lande hoort men hetzelfde, dus zou hij graag aan willen tonen dat zijn idee voor het hele land geldigheid heeft, zodat er wat aan gedaan kan worden. De manier om dat te doen is het uitspreken van een voorspelling, die hij uit zijn 'theorie' afleidt.

Deze voorspelling zou kunnen luiden: als het percentage leerlingen onder het verplegend personeel toeneemt, zal ook de ontevredenheid binnen het ziekenhuis toenemen. Over het meetbaar maken van de variabelen zullen we hier kort zijn. Het percentage leerlingen spreekt voor zichzelf en ontevredenheid besluit hij te gaan meten met een 5-puntsschaaltje (zie schaalconstructie, hoofdstuk 6) zeer tevreden - tevreden - weet niet - ontevreden - zeer ontevreden.

### *Gegevens verzamelen*

Het eerste cijfer kan hij vermoedelijk wel vinden in de boeken van de personeelsafdeling (inhoudsanalyse, hoofdstuk 5). Het vragen-

lijstje kan hij onder de proefpersonen verdelen, waarna zij dat zelf in kunnen vullen (schriftelijke enquête). Uiteraard had hij het ook alle mensen kunnen gaan vragen (vraaggesprek, hoofdstuk 3), of kunnen gaan observeren (hoofdstuk 2) hoe vaak er op elkaar gemopperd wordt enz.

Deze en nog andere vormen van gegevensverzameling komen uitgebreid in dit boek aan de orde.

### *Het ordenen en samenvatten van de gegevens*

In feite is dit boek nu uit, maar voor het onderzoek volgen nog een aantal zeer belangrijke fasen. Allereerst dient men de 'emmer' met getalletjes die men verzameld heeft te ordenen. Hiertoe kan men ze op allerlei manieren in tabellen en figuren zetten. Deze bezigheid wordt heel vaak onderschat en in veel publikaties worden deze resultaten zelfs weggelaten. Er wordt dan volstaan met het geven van samenvattende getallen zoals gemiddelde, samenhang of verschil. In zeer veel gevallen zijn die getallen echter niet voldoende om een redelijk inzicht te krijgen. Een handvol even grote knikkers weegt gemiddeld evenveel als een mespuntje zandkorrels gevoegd bij de Amersfoortse kei. In een tabel of een plaatje zou men dat meteen zien.

Omdat men naast het geven van een overzicht van de gegevens ook tot een uitspraak wil komen over het al of niet aanwezig zijn van samenhang (correlatie) of verschil, worden de gegevens samengevat in één getal, de correlatiecoëfficiënt of het verschil.

Vervolgens gaat men na of deze getallen wel iets voorstellen. Men kan dat op twee manieren doen, namelijk kijken naar de *significantie* of naar de *relevantie*. Het eerste houdt in dat men kijkt of dat getal door toeval kan zijn ontstaan. Het kijken naar relevantie houdt in dat men kijkt of het gevonden verschil of de samenhang qua grootte van belang is binnen de theorie. (Uiteraard kan het voorkomen dat een zeer klein verschil al relevant is, dan dient men ook de significantie te controleren, maar in verreweg de meeste gevallen zal relevantie veel zwaardere eisen stellen aan het materiaal dan significantie.)

Onze onderzoeker denkt dat het percentage leerlingen een zeer belangrijke invloed heeft op de sfeer. Een correlatie van .30 (een correlatie kan liggen tussen  $-1$  en  $+1$ ) zou hem dan ook zeer teleurstellen. Hij hoopt op minstens .60. Hoewel .30 zeker significant kan zijn (dat hangt hoofdzakelijk af van het aantal ziekenhui-

zen dat hij onderzoekt), zal hij dat niet als steun zien voor zijn theorie.

### *Wat heeft het onderzoek opgeleverd voor de theorie?*

In feite lijkt deze fase weer sterk op de eerste. Er zijn nieuwe feiten bijgekomen en men dient te bekijken of de eerder gevormde mening hierdoor versterkt wordt of dat men juist iets nieuws zal moeten verzinnen. Gesteld dat de reeds veel besproken onderzoeker een correlatie van .30 heeft gevonden, dan zal hij van het idee af moeten stappen dat al die leerlingen de sfeer bedorven hebben. We raden hem overigens aan eerst wat meer werk te maken van de literatuurstudie en zijn blinde darm nog maar even te laten zitten.

### *Samenvatting*

In het kort komt een onderzoek dus neer op de volgende handelingen:

1. Men verbaast zich over een bepaald verschijnsel.
2. Men bedenkt en zoekt naar variabelen en verbanden tussen deze variabelen die het verbazingwekkende verschijnsel kunnen verklaren.
3. Uit deze verklaring leidt men voorspellingen af.
4. Vervolgens verzamelt men gegevens waaruit men kan aflezen of die voorspellingen juist waren.
5. Men past een aantal 'trucs' toe op de gegevens zodat dat aflezen mogelijk wordt.
6. Als de voorspellingen zijn uitgekomen, kan men ofwel nieuwe voorspellingen doen, die een ander facet van de theorie belichten, ofwel proberen de theorie in daden om te zetten. Werden de voorspellingen weerlegd door de gegevens, dan dient men van voor af aan te beginnen en een nieuwe theorie te bedenken of de bestaande te wijzigen.

Het is duidelijk dat in al deze fasen fouten gemaakt kunnen worden. Voor de beoordeling van onderzoek is het nodig dat men inzicht in al deze schakels heeft en dat men leert over de gang van zaken te argumenteren.

## Aanbevolen literatuur

Delen van de hieronder genoemde literatuur zullen u bekend voorkomen.

D. J. Fox, *The research process in education*. New York, Holt, Rinehart & Winston, 1969.

A. D. de Groot, *Methodologie*. Den Haag, Mouton, 1972.

## 2. OBSERVATIE

Een verkenning van een sociaal-wetenschappelijke onderzoekstechniek

### 1. Inleiding

Observeren is vrijwel zeker de oudste methode van wetenschappelijk onderzoek. Men zou kunnen zeggen dat het gaat om een oervorm van wetenschappelijke informatieverwerving. Wetenschappen als biologie, natuurkunde, astronomie of geneeskunde zouden ondenkbaar zijn zonder de gegevens die gedurende vele eeuwen via observaties werden verzameld. Ook in de sociale wetenschappen is de observatie niet meer weg te denken. Strikt genomen kunnen alle dataverzamelingstechnieken observatie genoemd worden. De term observatie in deze brede betekenis kan in principe worden gebruikt voor methoden als vraaggesprek, vragenlijst en test. In dit hoofdstuk wordt observatie echter in een specifiekere betekenis gebruikt, d.w.z. als één van de methoden van sociaal-wetenschappelijk onderzoek. Hierbij wordt het accent gelegd op observatiemethoden die bruikbaar zijn voor het observeren van wat mensen doen (bijvoorbeeld in school- of gezinssituaties), dat voor de observator hier en nu waarneembaar is en dat al dan niet kan zijn vastgelegd op film of video.

Aanvankelijk werden slechts de natuurwetenschappelijke technieken (en idealen!) gebruikt bij de observatie van de sociale werkelijkheid. De resultaten van dit type onderzoek waren echter triviaal en teleurstellend. Als reactie hierop werd een aanval gedaan op het zonder meer toepassen van de 'exacte' benadering binnen de sociale wetenschappen, met name binnen de psychologie. Het gevolg was dat de *gestructureerde observatiemethoden* gedurende enige tijd uit de sociale wetenschappen werden verbannen, waarmee (zoals we nu weten) het kind met het badwater werd weggegooid. Tezelfder tijd vond er een opleving en herwaardering plaats van geesteswetenschappelijke benaderingen zoals introspectie en *verstehen*.

In de laatste tientallen jaren is er echter sprake geweest van een duidelijke vooruitgang in de ontwikkeling van bruikbare, gestructureerde observatiemethoden in de sociale wetenschappen. De minder gestructureerde (of 'vrije') methoden hadden aanvankelijk, nu op hún beurt, te lijden onder de herontdekking van de gestructureer-

de benadering. Hoewel een zekere wrijving tussen de aanhangers van beide benaderingen zeker niet verdwenen is, komt men steeds meer tot het inzicht dat de keuze van de methode in sterke mate afhankelijk dient te zijn van de aard van de vraagstelling (wat wil men door middel van observatie te weten komen), de fase waarin het onderzoek zich bevindt (bijvoorbeeld toetsend of exploratief) en de (meet-)situatie (de omgeving waarin de observatie plaatsvindt). Het is dus zeker niet juist dat het altijd beter is de gestructureerde methode te hanteren en dat de vrije, ongestructureerde methoden tweede keus zijn of slechts in vooronderzoek, bij het opwerpen van hypothesen, adequaat kunnen worden toegepast. Het kan uitermate vruchtbaar zijn beide benaderingswijzen te combineren bij de observatie van die aspecten van de sociale werkelijkheid die men onderzoekt.

## 2. *Observatie als methode*

### 2.1. *Observeren en waarnemen*

Observeren ligt, als je het oppervlakkig bekijkt, niet zover af van het begrip *waarnemen*. Bij het waarnemen, bijvoorbeeld kijken, luisteren, voelen en ruiken, worden verschijnselen door middel van de zintuigen in onze geest opgenomen. Deze zintuiglijke waarneming kan bewust of onbewust verlopen. Observeren onderscheidt zich van de menselijke waarneming door de *mate van bewustzijn*. Observeren is bewuste, doelgerichte waarneming. Kruijer<sup>1</sup> definieert observeren als 'het zorgvuldig en aandachtig gadeslaan of systematisch onderzoeken van verschijnselen met als doel de verschijnselen zo nauwkeurig mogelijk te leren kennen'. Ook De Groot<sup>2</sup> maakt een onderscheid tussen waarnemen en observeren. Waarnemen slaat op het ervaringsproces zonder er verder bij na te denken, terwijl observatie betrekking heeft op een meer systematische, doelgerichte houding van de onderzoeker.

### 2.2. *Observeren en denken*

Observatie veronderstelt een grotere afstand tot de zintuiglijke werkelijkheid dan de naïeve alledaagse waarneming, zoals we hierboven constateerden. Observatie steunt dan ook in sterkere mate op denken en theorievorming dan bij waarneming het geval is. Bij de observatie speelt niet alleen de theorie over het object van onderzoek een belangrijke rol, maar ook het waarnemen en observeren zelf wordt bij de theorievorming betrokken.



De waarneming wordt namelijk sterk beïnvloed door het denken, de ervaring, de taal en de behoeften van de mens en dat kan als men er niet op bedacht is bij de observatiemethode tot een vertekend beeld leiden.

Evenals bij de waarneming geldt bij de observatie dat het een activiteit is waarbij een beeld van de werkelijkheid wordt gemaakt. Bij de observatie echter gebeurt dit bewust en opzettelijk (in tegenstelling tot de alledaagse waarneming). Op basis van het voorgaande kan men concluderen dat iemand die over meer relevante kennis beschikt dan een ander een betere observator kan zijn. Iemand die meer weet over interactieprocessen tussen ouders en kinderen (op grond van ervaring en/of theorievorming) kan meer zien dan een oningewijde. Kruijer<sup>3</sup> citeert in dit verband een uitspraak van Loewenberg: 'The acquisition of 'fact' is not the reward of patient and careful observation. Facts are made, they are the products of intellectual effort.' De invloed van de gehanteerde theorie op de observatie wordt goed geïllustreerd in een voorbeeld afkomstig van Popper die het beeld van een zoeklicht gebruikt. De theorie strijkt als een zoeklicht over de werkelijkheid. Wát als feit zal kunnen verschijnen, is afhankelijk van de richting van de bundel, de intensiteit en de aard van de straling en de aanwezigheid van obstakels in de omgeving. Wat onze 'feiten' zullen zijn, hangt dus onder andere af van de inhoud van onze zoeklichttheorieën (Koningsveld<sup>4</sup>).

Men moet overigens niet in de (ernstige) fout vervallen om ter bevestiging van de juistheid van een theorie alleen dié observaties te *selecteren*, die goed bij de theorie aansluiten. Wetenschappelijk onderzoek is met een dergelijke handelwijze zeker niet gediend.

### 2.3. *Observeren en interpreteren*

In de alledaagse waarneming worden de verschillende aspecten van het waarnemingsproces niet bewust onderscheiden. Waarnemen, registreren, analyseren en interpreteren worden als het ware in één adem uitgevoerd zonder een vooropgezet plan of schema. De alledaagse waarneming biedt naast een aantal voordelen, zoals natuurlijkheid, betrokkenheid, doeltreffendheid, snelheid, tevens een aantal nadelen. Wie wel eens samen met anderen getuige is geweest van een vechtpartij of een verkeersongeval kan hierover meepraten. Vaak blijkt er slechts een fragment uit het totale gebeuren te worden waargenomen, terwijl waarneming en interpretatie niet worden onderscheiden. Behalve interpretatief is de waarneming dus ook dikwijls selectief. Zoals reeds eerder werd gesteld, hangt hetgeen men

in stimuli ziet tot op zekere hoogte af van de persoonlijke ervaring en geaardheid van de waarnemer.

In theorie is men in het algemeen zeker geneigd om bovenstaande opmerkingen met betrekking tot de subjectiviteit van het waarnemingsproces onmiddellijk te onderschrijven. In de praktijk heeft men echter toch de neiging aan (vooral de eigen!) observaties een objectief karakter toe te kennen. Een en ander kan worden geïllustreerd aan de hand van een aantal observatieverslagen dat door verschillende observatoren met betrekking tot een zelfde filmfragment werd vervaardigd. Cronbach<sup>5</sup>, van wie dit voorbeeld afkomstig is, vertoonde tweemaal een film (zonder geluid) over een jongen, met beelden over zijn gedrag in de klas en op de speelplaats en gaf de observatoren de opdracht louter weer te geven wat er getoond werd (interpretaties waren wel toegestaan, maar moesten in het verslag tussen haakjes worden geplaatst). Hier volgen, vrij vertaald, drie van de verslagen (nummers tussen haakjes corresponderen met scènes uit de film):

*Observator I:*

(2) Robert leest woord-voor-woord en gebruikt zijn vinger om de regel te volgen.

(4) Observeert meisje in box met sterke preoccupatie.

(5) Gedurende het zingen participeert hij over het algemeen niet erg actief. Belangstelling is voor een gedeelte van de tijd op iets anders gericht. Schijnt het meest actief te reageren bij delen van liedjes waar actie aan te pas komt. Heeft een neiging tot schijnbaar zinloze bewegingen. Draait met zijn vingers, zwaait doelloos met zijn armen.

*Observator II:*

(1) Glimlacht naar de camera (nieuwsgierig). Wanneer de groep opbreekt, maakt hij nerveuze gebaren, gooit zijn armen in de lucht.

(2) Aandacht voor de leesles. Leest met een ernstige blik. Moet een regelaanwijzer gebruiken.

(3) Zit meisjes achterna, plaagt.

(4) Meisje schopt wanneer hij zijn hand op haar been legt. Robert trekt een lelijk gezicht naar haar.

(5) Is aan het zingen. Zit met zijn mond open, slaat zijn knieën tegen elkaar, krabt aan zijn been, doet zijn vingers in zijn mond (schijnt verschillende nerveuze gewoonten te hebben, hoewel hij niet emotioneel verwrongen of te veel van zichzelf bewust is).

- (6) Twist met anderen. Verdedigt zijn recht.
- (7) Korte woordentwist, want hij wil op de ladder klimmen bij junglegymnastiek.

#### *Observator III:*

- (2) Gebruikt een regelaanwijzer om de woorden te volgen, leest langzaam, nogal geforceerd en met zorgvuldige vorming van geluiden (misschien onzeker van zichzelf en bevreesd fouten te maken).
- (3) Misschien lichtelijk agressief getuige het feit dat hij een jonger kind opzij duwt terwijl hij naar een andere plaats gaat. Speelt met andere kinderen met kennelijk plezier, glimlacht, rent, schijnt vooral in meisjes geïnteresseerd. Dit kan men opmerken bij spelletjes en ook bij het zingen.
- (5) Heeft weinig belangstelling voor het zingen, zit te wiebelen, beweegt handen en benen (misschien verlegen en nerveus). Schijnt bij liedjes onbekend te zijn met de woorden van de coupletten en toont ongeïnteresseerdheid door wiebelen en ronddraaien. Krijgt pas weer belangstelling bij het refrein. Zijn speciale kameraad schijnt een meisje te zijn, daar hij altijd bij haar zit.

Het is niet mogelijk uit bovenstaande verslagen precies op te maken wat er nu eigenlijk op het filmfragment te zien was. Men kan stellen dat de observatieverslagen niet alleen informatie verschaffen over het betreffende filmfragment, *maar ook over de observatoren*. Zij verschillen sterk in volledigheid. Lang niet alle scènes uit de film worden door alle drie gerapporteerd. Binnen bepaalde scènes worden verschillende aspecten vermeld. Hoewel de opdracht de mogelijkheid bood tot interpretatie (tussen haakjes), maken de observatoren nauwelijks onderscheid tussen observatie en interpretatie. Dit voorbeeld laat duidelijk zien dat waarneming selectief en interpretatief is en dat de observatoren elk vanuit een eigen referentiekader naar de film hebben gekeken.

#### 2.4. *Slotsom*

In de voorgaande paragrafen hebben we kunnen zien dat de observator zelf geen exact opname- en registratieapparaat is; zijn specifieke selectie-mechanismen kunnen de objectiviteit van de onderzoeksresultaten aanzienlijk beïnvloeden. Het is de vraag of het mogelijk en zinvol is te streven naar een absoluut onderscheid tussen naïeve waarneming en meer systematische observatie. Realistischer lijkt het te spreken van een overgang tussen beide vormen,

waarbij de naïeve waarneming geleidelijk moet worden vervangen door meer systematische en gecontroleerde procedures indien men streeft naar objectieve resultaten.

Samenvattend kan gesteld worden (en dit standpunt zegt iets over de selectie-mechanismen en referentiekaders van de auteurs!) dat observatie als een *wetenschappelijke onderzoeksmethode* kan worden beschouwd indien de volgende kenmerken aanwezig zijn:

1. de observatie dient een omschreven onderzoeksdoel, wordt uitgevoerd binnen het kader van een theoretisch raamwerk, m.a.w. is doelgericht;
2. er is sprake van een systematische planning;
3. de registratie van bijvoorbeeld interactie(s) tussen ouders en kinderen geschiedt systematisch (d.w.z. niet alleen datgene wat je bevalt of wat mooi in de theorie past);
4. de verkregen informatie dient valide en betrouwbaar te zijn en op dat punt controleerbaar (zie ook hfdst. 4).

De rol van de observator in het onderzoeksproces, met name ook de mate waarin men ernaar streeft hem uit te schakelen als foutenbron, wordt vaak gehanteerd als criterium voor het onderscheid tussen verschillende typen observatiemethoden. In de volgende paragrafen zullen wij hierop nader ingaan. Tevens zal aandacht besteed worden aan de meest gebruikte observatietechnieken.

### 3. *Observatievormen*

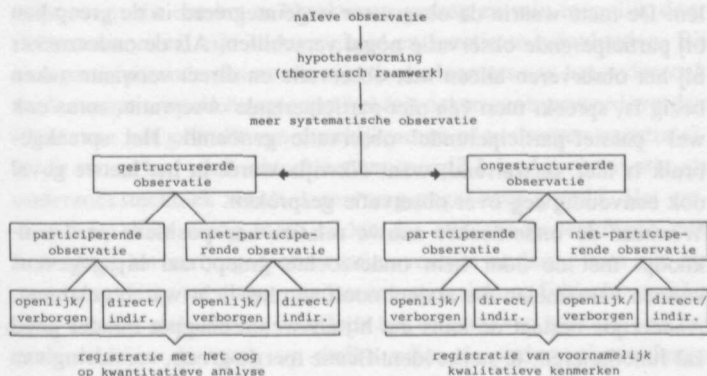
De indeling in verschillende observatievormen die hier wordt gepresenteerd is afkomstig van Sumaski<sup>6</sup> en bevat de meest voorkomende vormen. In het bovenste deel van het schema op blz. 31 vindt men nog eenmaal de overgang tussen naïeve en systematische observatie terug. Hieruit wordt duidelijk dat naïeve observatie kan leiden tot het formuleren van empirisch toetsbare hypothesen die ten grondslag liggen aan de toepassing van meer systematische observatietechnieken.

Aan de hand van dit schematisch overzicht gaan we nu kort in op de verschillende vormen.

#### 3.1. *Gestructureerde en ongestructureerde observatie*

In de eerste plaats dient opgemerkt te worden dat 'ongestructureerd' absoluut niet wil zeggen 'onwetenschappelijk'! Ongestructureerd observeren betekent: met een relatief wijdmazig net van observatiegezichtspunten en -categorieën meer *kwalitatief* gericht onderzoek

### Schematisch overzicht van observatievormen



doen op basis van een in algemene termen gesteld theoretisch kader. De observator verkent het onderzoeksgebied met behulp van algemene, vrij grove categorieën. Ongestructureerde observatie kan op deze manier fungeren als eerste stap op weg naar gestructureerde observatieprocedures. De ongestructureerde observatievorm kan echter ook in latere fasen van het onderzoek een rol blijven spelen, eventueel in combinatie met meer gestructureerde vormen.

Bij volledig gestructureerde observatie zijn er voor de te observeren variabelen nauwkeurige operationele definities voorhanden, die vervolgens op systematische wijze worden toegepast op de onderzoekselementen in een eveneens precies omschreven meetsituatie. Er ligt dus strikt vast *hoe* en *wat* waargenomen moet worden. Bij deze observatievorm is de rol van de observator aan strikte voorschriften gebonden en tot een minimum gereduceerd. *Kwantitatieve analyse van de gegevens is een belangrijke doelstelling.*

Het is van belang er op te wijzen dat gestructureerde en ongestructureerde observatie onzes inziens eerder gradueel dan principieel van elkaar verschillen. Men kan zich daarbij een continuüm voorstellen dat van relatief ongestructureerd, via allerlei tussenvormen, loopt naar relatief gestructureerd.

### 3.2. Participerende en niet-participerende observatie

Met participerende observatie wordt bedoeld dat de onderzoeker een groep observeert vanuit een positie die hij *in de groep* heeft ingenomen. Met andere woorden, hij neemt deel aan de activiteiten

van de groep die hij bestudeert door in die groep een rol te vervullen. De mate waarin de observator is geïntegreerd in de groep kan bij participerende observatie nogal verschillen. Als de onderzoeker bij het observeren alleen met observatie en direct verwante zaken bezig is, spreekt men van niet-participerende observatie, soms ook wel 'passief-participerende' observatie genoemd. Het spraakgebruik is hier verwarrend, want dikwijls wordt in het laatste geval ook eenvoudigweg over observatie gesproken.

Wanneer de onderzoeker nauwe relaties van positieve aard aanknoopt met de door hem onderzochte groep, zal hij gegevens kunnen verzamelen die anders nooit aan het licht waren gekomen. Anderzijds bestaat de kans dat hij als wetenschapper minder goed zal functioneren: te sterke identificatie met de groep, vervalsing van informatie, reconstructie achteraf. Als daarentegen de sociale afstand tot de groep te groot is, blijven bepaalde gegevens onbereikbaar.

Zo waren bijvoorbeeld de observaties van Thrasher<sup>7</sup> over de sociale structuur en het gedrag van bendes die nooit een observator in hun midden geduld zouden hebben, alleen realiseerbaar doordat de onderzoeker zelf lid werd van een groep uit de onderwereld. Wil participerende observatie in dit soort gevallen effectief zijn, dan moet de onderzoeker niet alleen een rol vervullen in de onderzochte groep, hij moet door de andere individuen ook worden geaccepteerd.

Bij niet-participerende observatie vervult de observator eigenlijk de rol van neutrale toeschouwer. Hij observeert een groep op afstand, of zit bijvoorbeeld achter in een schoolklas en poogt zich onzichtbaar te maken; desnoods houdt hij zijn plas op tot aan het speelkwartier.

Bij het bekijken van groepsprocessen via een gesloten t.v.-circuit of doorkijkramen (*one-way screens*) is de observatie extreem neutraal. Natuurlijk is ook hier het onderscheid tussen wel of niet participerende observatie niet absoluut en zijn allerlei tussenvormen mogelijk, maar in de praktijk van het onderzoek zal de keuze (van wel of niet participerende observatie) uiteindelijk toch afhangen van het doel van het onderzoek. Dat geldt natuurlijk ook voor gestructureerde of ongestructureerde observatie. Het geldt in feite voor alle andere observatievormen die nog aan de orde zullen komen.

### 3.3. *Openlijke en verborgen observatie*

Bij participerende observatie zal de onderzoeker zich dikwijls onop-

vallend proberen te gedragen en zijn eigenlijke bedoelingen niet prijsgeven; dit om het te observeren gedrag zo min mogelijk door (informatie over) zijn aanwezigheid en functie te beïnvloeden. Bij niet-participerende observatie zijn de deelnemers aan het onderzoek vaak evenmin op de hoogte van het feit dat ze geobserveerd worden (denk aan het gebruik van doorkijkramen of verborgen camera's). Is dit nu wel of niet *ethisch verantwoord*? Juist bij observatie als onderzoekstechniek wordt deze vraag nogal eens gesteld. Het antwoord op deze vraag moet luiden dat ethische bedenkingen met betrekking tot verborgen observatie evenmin op hun plaats zijn als met betrekking tot andere onderzoekstechnieken. Ethische problemen kunnen zich slechts voordoen ten aanzien van het doel van het onderzoek en de vraag op welke wijze de onderzoeksresultaten zullen worden gebruikt.

Bij openlijke observatie is de onderzoeker als zodanig bij de onderzochten bekend. In dit geval dient dus zeker rekening te worden gehouden met observator-effecten. Daarom is het begrijpelijk dat ook in het geval van openlijke observatie het eigenlijke doel van het onderzoek meestal niet (volledig) wordt meegedeeld. Wanneer een onderzoeker geïnteresseerd is in de mate waarin een groep werknemers bij het vervullen van hun taak afwijkt van de officiële bedrijfsvoorschriften zal hij dit niet meedelen, maar bijvoorbeeld zeggen dat hij onderzoek doet naar de 'taakverdeling en samenwerking in kleine groepen'.

#### 3.4. *Directe en indirecte observatie*

De term *indirecte observatie* verwijst naar technische registratievormen, met name cassette- of bandrecorder en film- of videoapparatuur. Met technische hulpmiddelen geregistreerde gebeurtenissen kunnen steeds opnieuw worden herhaald en zijn onontbeerlijk wanneer het van belang is exact aan te geven wat zich heeft afgespeeld. Wanneer kan het gebruik van video of film nuttig zijn (vgl. Bosch e.a.<sup>8</sup>):

1. indien de te observeren handelingen zeer snel worden voltrokken;
2. indien het te observeren gebeuren zo complex is dat het aandachtig bekijken van bepaalde aspecten ten koste gaat van andere essentiële aspecten;
3. bij het observeren van uiterst subtiele gedragsveranderingen;
4. bij het bestuderen van opeenvolgende veranderingen in (min of meer) complex gedrag;

5. indien het erom gaat zeer precieze metingen van bepaalde variabelen te verkrijgen die zeer kort of complex zijn;
6. bij het ontwikkelen van observatiecategorieën en het trainen van observatoren.

Het nadeel dat de camera(man) het gebeuren kan beïnvloeden kan enigszins worden ondervangen door gebruik te maken van een telelens of van een op afstand bedienbare camera.

#### 4. Systematische observatie

Een veelvuldig toegepaste vorm van gestructureerde observatie is de techniek der systematische observatie. Dikwijls wordt hierbij gebruik gemaakt van zogenaamde *observatiesystemen*. Een observatiesysteem kan worden omschreven als een geheel van handelingsvoorschriften en instrumenten voor het observeren, waarbij in principe alle onderdelen van het proces (observeren, coderen, analyseren en interpreteren) zodanig kunnen worden uitgevoerd dat herhaalbaarheid en controle mogelijk zijn. Observatiegegevens verzameld met deze methode zijn dikwijls geschikt voor nadere statistische analyse, terwijl wordt gestreefd naar generaliseerbare uitspraken. Bij systematische observatie beschikt de onderzoeker over een nauwkeurig gedefinieerd waarnemingsinstrument. Het waarnemingsdoel is expliciet; de variabelen van het onderzoek zijn reeds uitdrukkelijk geselecteerd in de vorm van begrippen en specifieke categorieën waarmee het te observeren gedrag kan worden beschreven. In het algemeen kan worden gesteld dat de observatie (binnen deze benadering) moet voldoen aan de methodologische eisen die gelden voor onderzoek volgens de *empirisch-analytische wetenschapstraditie* (objectiviteit, validiteit, betrouwbaarheid etc.).

##### 4.1. Het ontwerpen van observatiecategorieën

Zoals uit het voorafgaande kan worden afgeleid, is het voor het doelmatig toepassen van systematische observatie noodzakelijk dat een bepaalde selectie wordt gemaakt uit het totaal van feiten die voor de vraagstelling van de onderzoeker van belang kunnen zijn. Als men bijvoorbeeld geïnteresseerd is in het sociaal-emotionele klimaat binnen schoolklassen, dan zou de mate waarin een leerkracht openstaat voor initiatieven van de leerlingen een belangrijk onderdeel binnen het onderzoek kunnen zijn. 'De mate waarin een leerkracht openstaat voor initiatieven' is echter te vaag en ongeschikt om gehanteerd te worden *in het kader van systematische*



*observatie*. Het is nodig dit begrip te vertalen in bruikbare kenmerken die waarneembaar zijn. Dit betekent dat we het begrip nauwkeurig moeten definiëren en bepalen welke gedragsvormen zullen worden geobserveerd. Het begrip moet worden vertaald in concrete, waarneembare variabelen (operationalisatie). Als deze operationalisatie goed is gebleken ('valide', zie 5.2.) dan mogen we aannemen dat we meten wat we wilden meten. Is de onderzoeker zo ver dat hij (op grond van zijn theorie, op grond van resultaten uit ander onderzoek) weet waar hij in het gedrag op moet letten, dan kan hij aan de hand van die kennis een observatiesysteem ontwikkelen. Hij dient zich dan tevens te realiseren dat hij met zo'n systeem een selectie maakt uit alle mogelijke observeerbare gedragingen en tengevolge daarvan bezig is met een *gereduceerde werkelijkheid*, waarbij alleen dié aspecten van het gebeuren worden bestudeerd, die vanuit een bepaalde visie (theorie) relevant zijn.

Hierboven werd reeds meerdere malen gesproken over observatiesystemen. In de volgende paragraaf gaan we nader in op een aantal typen die vaak gebruikt worden.

#### 4.2. Soorten van observatiesystemen

Met betrekking tot de registratiewijze en/of manier van coderen van gebeurtenissen en daarmee samenhangende problemen is door een groot aantal auteurs een nog groter aantal onderscheidingen aangebracht. Hier zullen de onderscheidingen tussen de diverse observatiesystemen besproken worden die het meest gebruikt worden.

##### 4.2.1. Categorieënsystemen

De naam van dit type observatiesysteem is wat ongelukkig gekozen, omdat ook in andere typen wel degelijk gebruik wordt gemaakt van categorieën. Het kenmerkende van een categorieënsysteem is dat een bepaalde gebeurtenis (een bepaald gedrag) *iedere keer wanneer deze voorkomt* gescoord wordt in één van de categorieën. Bovendien is een categorieënsysteem (vanuit een bepaald gezichtspunt) *allesomvattend* en, in het ideale geval, zijn de ertoe behorende categorieën *wederzijds uitsluitend*.

De observator is in dit geval dus voortdurend aan het coderen. Kenmerken als allesomvattend en wederzijds uitsluitend worden zelden volledig gerealiseerd omdat (1) niet alle gebeurtenissen in categorieën te vatten zijn en omdat (2) vele gebeurtenissen bij twee of meer categorieën zijn onder te brengen (hetgeen resulteert in de noodzaak van lange trainingsperioden voor observatoren (*codeurs*))

en aanzienlijk minder dan 100% overeenstemming tussen de observatoren die het systeem gebruiken). Het gaat ook hier om het vinden van het juiste midden tussen de ontwikkeling van een heel ingewikkeld systeem met een groot aantal categorieën die subtiële onderscheidingen mogelijk maken, én een systeem met weinig categorieën dat een vrij grof onderscheid maakt, maar aanzienlijk makkelijker te hanteren is. De meeste tot nu toe ontwikkelde systemen bewegen zich tussen beide extremen. Beroemde voorbeelden van categorieënsystemen zijn het interactie-analysesysteem van Bales<sup>9</sup> en het observatiesysteem van Flanders<sup>10</sup>.

Eerstgenoemd instrument telt 12 categorieën voor de registratie en analyse van sociale interactie in kleine groepen, waarbij het er volgens Bales niet toe doet wat de opdracht of het thema is waarmee de groep zich bezighoudt. Volgens Bales is iedere mogelijke handeling op eenduidige wijze in één van de 12 categorieën onder te brengen.

Het observatiesysteem van Flanders omvat 10 categorieën om de interactie tussen leerkracht en leerling vast te leggen (7 voor leerkrachtgedrag, 2 voor leerlinggedrag en 1 restcategorie). Het gaat hierbij om het bepalen van het sociaal-emotionele klimaat in de les. Elke gedraging wordt in één van deze 10 categorieën gescoord. Bij dit systeem gebeurt dit om de 3 seconden. Door deze werkwijze blijft de volgorde van de gebeurtenissen behouden en krijgt men ook informatie over de duur van de gebeurtenissen. Bij het registreren worden alleen de categorienummers opgeschreven.

Voorbeeld:

Categorie:	4	8	8	3	9	7	4	9	9	9	
			\ /					\ /		/	
			\ /					\ /		/	
Seconden:	3		6	3	3	3	3		9		(totaal 30)

#### 4.2.2. Tekensystemen

Tekensystemen (*sign systems*) worden gebruikt wanneer het erom gaat het optreden van een of meer bepaalde gebeurtenissen vast te stellen. Daarbij is tevoren vastgelegd welke gebeurtenissen geregistreerd worden. Vaak wordt gebruik gemaakt van een lijst van gebeurtenissen (gedragingen) die in een bepaalde situatie kunnen voorkomen. Aan de hand van zo'n lijst wordt bekeken of een bepaalde gebeurtenis gedurende een van te voren gekozen tijdsinterval heeft plaatsgevonden. Binnen dit tijdsinterval wordt één soort gedrag, ongeacht de frequentie, slechts één keer gescoord.

Een voorbeeld:

Tijd : 9.00 uur Activiteit : werken met speelleermateriaal Leeftijdsgroep: 5-jarige kleuters										
Onzelfstandigheid	Interval (1 minuut)									
	1	2	3	4	5	6	7	8	9	enz.
Vraagt om hulp										
Zoekt lichamelijk contact										
Zoekt toenadering										
Vraagt om aandacht										

Met bovenstaand eenvoudig tekensysteem wordt elke minuut aangekruist of één van de vier gedragingen is voorgekomen. Hierbij kijkt men één minuut, kruist aan, kijkt weer één minuut, kruist aan enz.; of het gedrag 4 of 5 keer per minuut is voorgekomen, is niet van belang, het gaat er alleen maar om of het zich wel of niet voordoet.

#### 4.2.3. Beoordelingsschalen

Het ontwerpen van goede en bruikbare observatiesystemen vergt nogal wat tijd. Als men daarover niet in ruime mate beschikt of als het gaat om een meer globale indruk, kan men gebruik maken van beoordelingsschalen (*rating-scales*). Deze nemen meestal de vorm aan van een 5- of 7-puntsschaal: een gelijkmatig verdeelde continue lijn met aan de uiteinden twee uiterste waarden (zie hfdst. 7). De observator of beoordelaar geeft op een schaal aan in welke mate een bepaald kenmerk aanwezig is, hoe intens een bepaald gedrag zich manifesteert. Het gaat hierbij dikwijls om een mening of indruk van het waargenomene. De gekozen gedragingen of aspecten zijn nogal eens vaag en globaal omschreven, waardoor er meer ruimte is voor interpretatie dan in het geval van de nauwkeurig gedefinieerde categorieën van de eerder besproken systemen. De betrouwbaarheid van beoordelingsschalen geeft, gegeven het subjectieve karakter, vrij vaak problemen. De winst aan tijd in vergelijking met de constructie van een categorieënsysteem gaat dus vaak wel gepaard met een verlies aan nauwkeurigheid.

Beoordelingsschalen worden dikwijls gebruikt om gebeurtenissen te scoren *nadat* deze hebben plaatsgevonden. Ze kunnen echter ook tijdens het observeren in combinatie met een categorieënsysteem gebruikt worden om bijvoorbeeld de intensiteit of kwaliteit van een bepaald gedrag aan te geven. Zo zit in categorie 7 van het Flanders-systeem elke vorm van kritiek, van een woedeuitbarsting tot een

zakelijke verwerping. Door het gebruik van een beoordelingsschaal naast het observatiesysteem kan nu ook het karakter van de kritiek worden aangegeven.

Beruchte fouten die kunnen voorkomen bij het hanteren van beoordelingsschalen zijn het systematisch te hard of te vriendelijk oordelen, voornamelijk in het midden óf juist aan de uiteinden van de schaal scores (onafhankelijk van het vertoonde gedrag) en het optreden van het *halo-effect*, d.w.z. de neiging om bij de beoordeling van een bepaald kenmerk van een persoon beïnvloed te worden door een ander kenmerk of de totaalindruk van die persoon.

#### 4.3. *Keuze van analyse-eenheden: time sampling en event sampling*

Tot nu toe werd met betrekking tot de registratiewijze vooral gesproken over de structuur van de meest gebruikte observatiesystemen. Je kunt je echter ook afvragen hoe de stroom van gebeurtenissen of gedragingen uiteengerafeld moet worden terwille van de analyse. Deze invalshoek richt zich dus meer op de verschijnselen die in aanmerking komen voor observatie, *de analyse-eenheden*. Het meest frequent gemaakte onderscheid betreffende de keuze van analyse-eenheden is ongetwijfeld dat tussen *time sampling* en *event sampling*.

De *time sampling*-methode vestigt de aandacht op geselecteerde aspecten van een gebeurtenis of gedragsproces zoals dat optreedt binnen gelijkvormige, kortdurende intervallen. Het te observeren gedrag moet vrij vaak voorkomen, anders heeft *time sampling* geen zin. Er is veel variatie in de tijdsintervallen die in onderzoek gebruikt zijn. Perioden van vijf minuten of minder worden het meest gebruikt. De methode houdt nauwelijks rekening met natuurlijke gedragseenheden: aan het eind van het tijdsinterval stopt de observatie, ook al is de gedragsketen als zodanig nog niet afgelopen. Er wordt naar gestreefd om de observatie-segmenten zo te kiezen dat ze zoveel mogelijk kenmerkend zijn voor te observeren situaties in het algemeen. Voordelen van *time sampling* zijn onder meer de mogelijkheid tot systematische controle, betrouwbaarheid en economisch gebruik van onderzoekstijd en -energie.

*Time sampling* zou je kunnen hanteren bij de observatie van het gedrag van een leerkracht voor de klas. Er worden bijvoorbeeld drie perioden van vijf minuten in een lesuur gekozen (in het begin, in het midden en tegen het eind van de les) en vergelijkbare perioden tijdens het eerste, het derde en het vijfde lesuur. Als je het heel goed wilt doen, zorg je ervoor dat de steekproeven (*time samples*) niet

alleen in het begin van de week, maar ook in het midden en tegen het einde van de week worden getrokken. Het gaat er bij deze methode immers om dat de tijdstippen zodanig worden gekozen, dat ze gezamenlijk als representatief kunnen worden beschouwd voor het te onderzoeken verschijnsel.

Naast time sampling wordt ook event sampling nogal eens toegepast. De observatie- of analyse-eenheid is in dit geval meer 'natuurlijk', in de zin van afgerond. Niet de tijd geeft het begin en einde van de eenheid aan, maar de gebeurtenis zelf. De duur van elke eenheid is dus flexibel. Het is de onderzoeker er om te doen bepaalde gebeurtenissen in hun geheel te bestuderen. De observator wacht tot het te observeren gedrag zich voordoet en beschrijft dan nauwkeurig hoe en wanneer het optreedt. De methode is ook bruikbaar bij gedragingen die slechts zelden optreden. Het gedrag kan bij gebruik van deze methode beter in zijn context worden bestudeerd dan met time sampling het geval is. Event sampling zou toegepast kunnen worden bij het bestuderen van ruzie(tjes) tussen peuters in een crèche. Zodra de observator waarneemt dat er ergens een ruzie tussen twee (of meer) kinderen dreigt te ontbranden, neemt hij de pen (of de cassetterecorder) ter hand en beschrijft zo nauwkeurig mogelijk de gang van zaken, of dit nu één minuut of tien minuten duurt. De 'event' was hier dus 'ruzie tussen kinderen'.

In het algemeen geldt dat de keuze van de analyse-eenheid afhankelijk is van de specifieke doelstellingen van de onderzoeker. Is de *vraagstelling* gericht op een bepaald gedrag en de relatie van dat gedrag tot de omgeving waarin het voorkomt, dan is event sampling een geschikte methode. Als de *vraagstelling* betrekking heeft op veel voorkomend gedrag, dan is time sampling een goede methode om op economische wijze gegevens te krijgen over een vrij lange periode en/of van een grote groep personen. Ook het observatiesysteem (categorieën-, tekensysteem, beoordelingsschaal) bepaalt vanzelfsprekend de keuze van een bepaalde analyse-eenheid. Het blijft daarom moeilijk algemene richtlijnen te geven.

## 5. *Betrouwbaarheid en validiteit van observatiegegevens*

### 5.1. *Betrouwbaarheid*

Het is van groot belang bij het gebruik van observatiegegevens (of van observatie *afgeleide* gegevens) in het onderzoek nauwkeurig aan te geven welke betrouwbaarheidsmaten men heeft gekozen en waarom. Binnen de literatuur over observatie wordt de term 'be-

trouwbaarheid' slordig gebezigd en door velen op een 'eigen wijze' geïnterpreteerd.

De 'betrouwbaarheid' van observaties wordt vaak op de volgende drie manieren gedefinieerd (en meestal berekend in de vorm van coëfficiënten):

A. Als de mate van intersubjectieve overeenstemming tussen twee of meer observatoren met betrekking tot hun coderingen van een zelfde situatie. Ze kunnen daartoe tegelijkertijd ergens aanwezig zijn en hun observaties verrichten of ze coderen onafhankelijk van elkaar een zelfde video-opname. Een andere benaming voor deze overeenstemmingsmaat is o.a. *intercoder agreement*. Het gaat bij deze maat met name om de *objectiviteit van het observatiesysteem als meetinstrument*: in welke mate is het waargenomen een kenmerk van het waargenomen object en geen toevallig kenmerk van de observator of een onzorgvuldig gedefinieerde categorie.

De gemakkelijkste manier om de intersubjectieve overeenstemming vast te stellen is de volgende. Een zelfde op video geregistreerd lesfragment wordt door twee observatoren gecodeerd. Vervolgens worden hun observaties met elkaar vergeleken. De overeenstemming tussen de observatoren kan geschat worden met de volgende formule:

$$\text{Overeenstemming} = 1 - \frac{A - B}{A + B}$$

De coderingen van de twee observatoren worden in de teller van elkaar afgetrokken en in de noemer bij elkaar opgeteld, waarbij het grootste getal steeds voorop staat. Bijvoorbeeld: observator A heeft 25 kritische opmerkingen van de leerkracht gecodeerd en observator B slechts 20. De overeenstemming tussen A en B bedraagt dan:

$$1 - \frac{25 - 20}{25 + 20} = 1 - 0.11 = 0.89$$

Een andere veel toegepaste formule ziet er zo uit:

$$\frac{\text{aantal overeenstemmingen}}{\text{aantal overeenstemmingen} + \text{aantal niet-overeenstemmingen}} \times 100$$

Het aantal overeenstemmingen heeft betrekking op het aantal scores dat beide observatoren gemeen hebben, terwijl het aantal niet-overeenstemmingen betrekking heeft op het aantal gedragingen dat de ene observator wel en de andere niet gecodeerd heeft.

Er zijn onderzoekers die vinden dat deze intersubjectieve overeenstemming minstens 0.85 (of 85%) moet bedragen om waarde te kunnen hechten aan de observatiegegevens. In het kader van exploratief onderzoek wordt de ondergrens dikwijls wat lager gesteld, bijvoorbeeld op 0.60 (of 60%).

B. Als de mate van *intrasubjectieve overeenstemming* (*intracoder agreement*): waarbij observatieresultaten van dezelfde observator die herhaalde malen hetzelfde beeld observeert, worden vergeleken. Deze overeenstemmingsmaat geeft informatie over de veranderlijkheid van de observator. Het gaat om de stabiliteit van de coderingen die de observator toekent aan (hetzelfde) gedrag. Men spreekt hier dan ook wel van de stabiliteitscoëfficiënt. Hoewel daar op bepaalde gronden bezwaar tegen kan worden gemaakt, gebruikt men voor het bepalen van de intrasubjectieve overeenstemming in de praktijk dikwijls de onder (A) geïntroduceerde formules.

C. Als de correlatie tussen coderingen van verschillende observatoren op verschillende tijden, in verschillende situaties, onder verschillende omstandigheden. Deze betrouwbaarheidsmaat staat bekend onder de naam *betrouwbaarheidscoëfficiënt* (*reliability coefficient*). Abstract uitgedrukt geeft deze coëfficiënt aan welk deel van de gevonden verschillen echt is. De eerder genoemde maten hielden alleen rekening met de vertekening die veroorzaakt wordt door het werken met niet perfecte observatoren. Deze coëfficiënt houdt ook rekening met andere invloeden en geeft dus een vollediger benadering van het betrouwbaarheidsprobleem.

De onder (A) en (B) genoemde overeenstemmingsmaten geven alleen informatie over het 'instrument', bijvoorbeeld de mate van ondubbelzinnigheid van de categorieën, de geoefendheid van de observatoren en dergelijke. Als je meer wilt weten over de *betrouwbaarheid van de gegevens* die je met behulp van het observatiesysteem hebt verkregen, d.w.z. in welke mate deze observatiegegevens zijn toe te schrijven aan *systematische variantiebronnen*, kom je met (C) in de buurt, maar is de techniek der variantie-analyse de beste keuze. Een voorbeeld kan dit duidelijk maken (zonder inhoudelijk in te gaan op wat variantie-analyse nu precies is): als je op basis van gegevens verkregen met behulp van een observatiesysteem een aantal leerkrachten zou willen vergelijken met betrekking tot hun onderwijsstijl, zijn er een aantal 'variantiebronnen' in het spel waar je rekening mee móet houden. De leerkrachten zelf op de eerste plaats, het leerjaar (gaat het om een eerste of om een vijfde

klas), het soort les (Engels of wiskunde), het tijdstip van observatie ('s morgens vroeg of in de namiddag), de observator (meneer A of meneer B) etc. Naarmate je meer weet van de invloed van elk van deze variantiebronnen op de score die het observatiesysteem oplevert, stijgt de *betrouwbaarheid van die score als maat voor het verschil tussen de leerkrachten*.

## 5.2. Validiteit

Bij validiteit gaat het om de mate waarin theoretische begrippen (zoals angst, vreugde, creativiteit e.d.) in de werkelijkheid zijn terug te vinden en het mogelijk maken die werkelijkheid te verklaren, én de mate waarin de 'vertaling' van die begrippen in de vorm van concrete, meetbare variabelen (*operationalisatie*) in zijn betekenis overeenstemt met de oorspronkelijke theoretische begrippen. Concreter geformuleerd: meten de observatiescores inderdaad wat ze beogen te meten? Dit kan men controleren indien men de beschikking heeft over een onafhankelijke meting van dezelfde variabele, waarmee de observatiescores dan vergeleken kunnen worden. Dergelijke metingen zijn echter zelden voorhanden, of beter gezegd, men heeft zelden moeite gedaan dergelijke onafhankelijke metingen te verrichten. Het onderzoek naar de validiteit van observatiegegevens is een nogal verwaarloosd gebied. De meeste ontwerpers van observatiesystemen komen niet verder dan de poging validiteit uit te drukken in de mate waarin observatoren overeenstemmen. Eerder hebben we gezien dat dit echter alleen iets zegt over de objectiviteit van het observatiesysteem als meetinstrument, hetgeen hoogstens een (bescheiden) aspect is van betrouwbaarheid. In navolging van de testpsychologie (waar men al zo'n jaar of 80 ervaring heeft met het validiteitsprobleem) kan men de validiteit van observatieresultaten op verschillende manieren bekijken:

- *predictieve validiteit*: kun je op basis van observatiegegevens het gedrag waarin je geïnteresseerd bent, *voorspellen*?
- *concurrente validiteit*: in hoeverre stemmen de observaties overeen met andere tegelijkertijd verkregen gegevens (zoals tests, rapporten e.d.)?
- *construct validiteit*: in hoeverre passen de observatieresultaten in bepaalde theoretische kaders?
- *inhoudsvaliditeit*: hoe goed representeert de observatie (het observatiesysteem) de gebeurtenissen of gedragingen die van belang zijn voor het gestelde doel?

Overigens kunnen betrouwbaarheid en validiteit niet los van elkaar



gezien worden. Betrouwbaarheid van een observatiesysteem is een noodzakelijke, maar niet voldoende voorwaarde voor de validiteit van de observatiescores. Een onbetrouwbaar systeem kan niet valide zijn, maar een betrouwbaar systeem is niet altijd valide. (Zie ook 6.1.)

## 6. *Systematische of 'vrije' observatie?*

In het voorafgaande is meerdere malen naar voren gekomen dat er verschillende observatievormen zijn (vgl. 3). In paragraaf 4 is met name gesproken over systematische observatie, een sterk gestructureerde observatievorm. Eén van de redenen dat juist aan deze vorm zoveel aandacht werd besteed is, dat het daarbij goed mogelijk is allerlei voorschriften, normen en regels, die bij het wetenschapsbedrijf van belang zijn, de revue te laten passeren. Bij de minder systematische, relatief ongestructureerde observatievormen (soms wordt hier gesproken over 'vrije' observatie) zijn er vooraf geen uniforme en strenge regels opgesteld. Getracht wordt juist om het te observeren object (de te observeren situatie) als een totaliteit te doorgronden, zonder van te voren uit te gaan van operationele begrippen, modellen of hypothesen. Bij de systematische benadering is het tegendeel het geval. Zoals eerder gesteld (vgl. 3.1) beschouwen wij gestructureerde en ongestructureerde observatievormen niet als principieel van elkaar verschillend, maar meer als een geleidelijke overgang die van relatief ongestructureerd, via allerlei tussenvormen, loopt naar relatief gestructureerd. Verder hebben wij ook opgemerkt dat de ene vorm niet beter is dan de andere, maar dat tal van overwegingen een rol spelen bij de keuze van de ene of de andere benadering (vgl. 1).

Wat zijn nu de voor- en nadelen van deze twee benaderingswijzen? We gaan daar hieronder op in.

### 6.1. *Voor- en nadelen van systematische observatie*

De systematische observatie, ingebed in de empirisch-analytische wetenschapstraditie, moet voldoen aan de methodologische eisen die binnen deze traditie gelden: controle, herhaalbaarheid en kritische evaluatie van de resultaten moeten mogelijk zijn. De observaties zijn niet gekoppeld aan de bijzondere kwaliteiten van de onderzoeker. De observator is niet een soort 'ziener' die dingen ziet en ontdekt waaraan gewone mensen voorbijgaan. De resultaten zijn niet alleen mededeelbaar, ze zijn ook herhaalbaar mits men zich aan

de regels houdt. Juist omdat de observatie berust op expliciete voorschriften en procedures is kritiek mogelijk en zinvol. Naast deze voordelen is er ook stevige kritiek te leveren op de systematische aanpak, en wel op de volgende punten:

– *scheiding observator en object*: tussen observator en object wordt een systeem van regels, procedures en categorieën geschoven. In plaats van dialoog en deelname aan het gebeuren is er distantie via een koel, technisch taalgebruik. De directe ervaring wordt opgegeven ten gunste van strikte toepassing van methodologische regels.

– *beperving tot uiterlijk gedrag*: alleen datgene wat gemeten en gecategoriseerd kan worden, speelt een rol in de observatie. Intenties, angsten, motieven, verwachtingen van mensen blijven buiten beschouwing (*black box, input-output model*).

– *reductie van de werkelijkheid*: zinvolle gedragsgehelen worden opgesplitst in nietszeggende stukjes. Uit het totale gedrag wordt slechts één onderdeel geobserveerd. Vaak wordt ook gebruik gemaakt van technische hulpmiddelen ten behoeve van registratie van het gebeuren in plaats van studie van de levende werkelijkheid. Dikwijls wordt ook de *context* van het gebeuren veronachtzaamd en wordt te veel aandacht geschonken aan het vinden van algemene wetmatigheden, los van de context.

– *te veel aandacht voor de groep*: gedragingen van afzonderlijke individuen of interacties tussen afzonderlijke individuen worden vaak niet in kaart gebracht.

Genoemde kritiek is met name afkomstig van kwalitatief georiënteerde onderzoekers, die een meer 'vrije', ongestructureerde observatievorm voorstaan. In de volgende paragraaf gaan we nader in op de voor- en nadelen van deze benadering.

## 6.2. Voor- en nadelen van 'vrije' observatie

Een kwalitatieve, 'vrije' observatievorm kan ten aanzien van de systematische aanpak zeker een aantal voordelen opleveren:

– complexe verschijnselen worden als *totaliteit* beschreven en geïnterpreteerd, zonder vooropgezette schema's en categorieën (levensechtheid en nauwgezetheid van de beschrijvingen).

– de vrije aanpak kan van grote betekenis zijn, vooral in een beschrijvende onderzoeksfase, voor het opsporen van categorieën en voor het ontwikkelen van theorieën die gefundeerd zijn in de sociale werkelijkheid.

– de observatie beperkt zich niet tot uiterlijk waarneembaar gedrag; innerlijke aspecten van het handelen van mensen (intenties, gevoe-

lens e.d.) worden als essentieel gezien.

– in plaats van de pretentie van generaliseerbaarheid wordt het contextgebonden karakter van de resultaten benadrukt.

– de resultaten zijn begrijpelijker voor niet-onderzoekers, i.c. de mensen die bij het onderzoek betrokken waren.

Naast deze voordelen is er, in vergelijking met de systematische aanpak, een aantal ernstige problemen (in de ogen van de 'systematici') die zich bij de vrijere benaderingen (kunnen) voordoen:

– het ontbreken van intersubjectieve controle en herhaalbaarheid. Een gevolg van het observeren zonder vooropgezette modellen en categorieën met betrekking tot het object.

– de observatoren maken gebruik van subjectieve, onuitgesproken categorieën, die in hoge mate bepalend zijn voor het observatiere-sultaat.

– het ontbreken van ondubbelzinnige, gedetailleerde observatie-procedures maakt herhaling en controle uiterst moeilijk. Betrouw-baarheid is nauwelijks te bepalen.

– bij kwalitatieve observatie wordt een zware wissel getrokken op het vakmanschap van de observator. Hierdoor is een scheiding tussen onderzoeker en dataverzamelaar (*observator*) vaak niet mo-gelijk. Dit maakt het extra moeilijk eigen vooronderstellingen, vooringenomenheid e.d. buiten spel te zetten.

Een zekere systematiek is ook bij deze benadering wel te realiseren. Middelen hiertoe zijn bijvoorbeeld: voortdurende vergelijking, het zoeken naar tegenvoorbeelden, het inzetten van meer observatoren, supervisie van de observator en het gebruik van protocollen voor het maken van notities van allerlei soort.

### *Aanbevolen literatuur*

Delen van de hieronder genoemde literatuur zullen u bekend voor-komen.

Terwel, J., 'Observeren in het onderwijs: Theorie, methoden en kritiek'. In: *INFO*, 9e jaargang, nummer 4, april 1978; Instituut voor Onderwijskunde, R.U. Groningen.

Sumaski, W., *Systematische Beobachtung: Grundlagen einer em-pirischen Methode*. Hildesheim, Georg Olms Verlag, 1977.

Kok, J. en S. Veenman, *Onderwijsleerpakket 'Interactie Analyse', deel 2a: Observeren in het onderwijs*. Den Bosch, Katholiek Peda-gogisch Centrum, Afdeling Opleidingen, 1977.

Friedrichs, J. & H. Lüdtkke, *Teilnehmende Beobachtung*. Beltz Verlag, Meinheim und Basel, 1977.

Hendriks, A. F. C. J., *Systematische gedragsobservatie bij het onderzoek van gezinsinteractie*. Akademisch proefschrift, Nijmegen, 1975.

Tavecchio, L. W. C., *Quantification of teaching behavior in physical education. A methodological study*. Dissertatie, Universiteit van Amsterdam, 1977.

### Noten

1. G. J. Kruijer, *Observeren en redeneren*. Meppel, Boom, 1973, blz. 13.
2. A. D. de Groot, *Methodologie*. Den Haag, Mouton, 1972, blz. 29.
3. a. w., blz. 14.
4. H. Koningsveld, *Het verschijnsel wetenschap: Een inleiding tot de wetenschapsfilosofie*. Meppel, Boom, 1973, blz. 101.
5. L. J. Cronbach, *Essentials of psychological testing*. Tokyo, Harper International Student Reprint, 1965.
6. W. Sumaski, *Systematische Beobachtung: Grundlagen einer empirischen Methode*. Hildesheim, Georg Olms Verlag, 1977, blz. 49.
7. F. M. Thrasher, *The gang: A study of 1.313 gangs in Chicago*. Chicago, 1968.
8. J. D. Bosch, J. S. Dijkstra, S. M. J. Hekken en H. Nakken, *Leren observeren: Een introductie in het gebruik van systematische gedragsobservaties*. Muiderberg, Coutinho, 1979, blz. 26.
9. R. F. Bales, *Interaction process analysis: A method for the study of small groups*. Reading, Mass., Addison-Wesley, 1950.
10. N. A. Flanders, *Analyzing teaching behavior*. Reading, Mass., Addison-Wesley, 1970.

### 3. HET VRAAGGESPREK

#### *Inleiding*

Een gesprek, óók onze dagelijkse conversatie, heeft altijd een doel. Soms is het gesprek een doel in zichzelf, gewoon voor de gezelligheid, meestal echter gaat het om het uitwisselen van informatie. Vaak is die ruil wederzijds, soms gaat de stroom alleen van de initiatiefnemer naar de ander, zoals bij voorlichting en dergelijke, in andere gevallen heeft de initiatiefnemer het doel kennis te vergaren. Die laatste vorm is onderwerp van dit hoofdstuk: gesprekken met het doel kennis 'af te tappen' van respondenten.

De kennis die verzameld wordt, kan verschillende zaken betreffen. Men kan iets willen weten over de persoon zelf (leeftijd of opleiding), iets over zijn ervaringen (ooggetuigeverslagen), iets over zijn kijk op bepaalde zaken (opiniepeilingen), over attitudes van de ondervraagde persoon (rassendiscriminatie), of iets over zijn gedrag in verleden of toekomst. Juist doordat men zoveel verschillende zaken kan onderzoeken door middel van het vraaggesprek (en iedereen het praten beheerst), is het vraaggesprek de dataverzamelmethode die binnen het wetenschappelijk onderzoek het meest gebruikt wordt, hoewel het vaak niet de enige bron van gegevens is, maar bijvoorbeeld wordt gebruikt om te controleren of een bepaalde experimentele conditie ervaren is zoals de onderzoeker dat bedoelde.

Hoewel dus vaak toegepast, is het zeker geen eenvoudige methode. Het aantal fouten dat gemaakt kan worden of gemaakt wordt, is vrijwel niet te overzien. Vandaar dat we eerst beginnen met enige voorwaarden waaraan een vraaggesprek moet voldoen om bruikbare gegevens op te kunnen leveren. (Let wel: als men aan deze voorwaarden voldoet, houdt dat niet in dat de gegevens ook juist zijn.) In de eerste plaats geldt natuurlijk weer, dat men precies moet bepalen wat men te weten wil komen. (Open interviews, literatuurstudie en nadenken kunnen hier weer bij helpen.) Dan dient men te bekijken of dat wel gerealiseerd kan worden door middel van het afnemen van een vraaggesprek. Hierbij zijn in ieder geval de volgende punten van belang:

1. De respondent moet beschikken over de gevraagde informatie. Om verschillende redenen kan de verlangde kennis afwezig zijn. In de eerste plaats kan het zijn dat de ondervraagde de informatie nooit heeft gehad. (Alleen in quizen schijnt het vermakelijk te zijn mensen te ondervragen over dingen die ze niet weten, hoewel sommige studenten over tentamens hetzelfde idee hebben.) In de tweede plaats kan men de vragen zo stellen dat de respondent zegt het niet te weten, terwijl hij wel over de kennis beschikt. (Veel ouders zullen bijvoorbeeld het antwoord schuldig blijven op de vraag: 'Heeft uw kind last van enuresis nocturna.')

In de derde plaats kan de kennis verdwenen zijn. Soms is men de informatie gewoon vergeten, in andere gevallen kan het een kwestie van niet meer willen weten (verdringing) zijn. In het eerste geval (vergeten) kan men via doorvragen misschien het geheugen van de respondent opfrissen; heeft men echter met verdringing te maken dan zal dat weinig helpen. Bij normaal vergeten zijn twee factoren van belang: de tijd die verlopen is sinds de gebeurtenis en het belang dat de respondent hecht aan die gebeurtenis. Met het eerste verschijnsel is iedereen wel op de hoogte. Hoe verder de gebeurtenis in het verleden ligt, hoe vager de herinnering. Het tweede verschijnsel is iets interessanter omdat hierdoor niet alleen de exactheid van de herinnering beïnvloed wordt, maar ook de richting van de fout. Het aantal leningen dat men bijvoorbeeld heeft lopen wordt, vooral onder de genietters van een wat hoger inkomen, veel te laag opgegeven, terwijl het aantal vliegtochten dat gemaakt is veel groter is dan in werkelijkheid, als men dit vaststelt door middel van een interview. Beide informaties worden als belangrijk gezien, de één blijkt echter minder en de ander meer acceptabel te zijn. Als tegelijk de tijd en het belang verschillen, kan dat elkaar versterken of verzwakken. Belangrijke gebeurtenissen uit een ver verleden worden beter onthouden dan onbelangrijke recente.

2. De respondent moet de situatie begrijpen.

Tijdens een interview wordt een heel specifieke rol verwacht van de respondent, namelijk dat hij antwoorden geeft en niet dat hij ook vragen gaat stellen. Daarnaast wordt verwacht dat hij op een bepaalde manier de vragen beantwoordt. Zo beantwoordde een aantal boeren de vraag of zij vruchtbomen bezaten ontkennend, terwijl later bleek dat een groot deel van hen wel degelijk vruchtbomen had, zij het slechts één of twee. Zij vatten het vraaggesprek blijkbaar op als een onderzoek naar hun inkomen, terwijl de interviewer andere dingen te weten wilde komen. Het is dus belangrijk dat de

respondent begrijpt wat de interviewer wil weten. Daarom moet men altijd in de inleiding of introductie iets zeggen over de inhoud en het doel van het interview.

3. De respondent moet gemotiveerd zijn de juiste antwoorden te geven.

In feite valt dit in twee delen uiteen: in de eerste plaats moeten ze gemotiveerd zijn *antwoord* te geven, in de tweede plaats moeten dat ook nog de *juiste antwoorden* zijn. Deze twee problemen zullen nooit geheel worden opgelost.

Tijdens vrijwel elk onderzoek waarbij interviews gebruikt worden, zijn er mensen die weigeren mee te doen. Dit kan om heel uiteenlopende redenen gebeuren. Een aantal mensen weigert elk interview. Daar valt weinig aan te doen. Anderen weigeren omdat zij bang zijn de antwoorden niet te weten. Dit kan deels ondervangen worden door de inhoud van het interview bekend te maken, te zeggen dat de antwoorden anoniem zullen blijven en, indien van toepassing op het vraaggesprek, dat het om een mening gaat en niet om hun kennis. Daarnaast zijn er wat trucjes, veelal ontwikkeld door maatschappijen die hun waren deur aan deur slijten, die enige ethische problemen op kunnen leveren. Deze variëren van het beloven van cadeautjes, de voet-tussen-de-deur-truc, het uiten van bedreigingen, tot de volslagen mislukkelingen-rol. Deze laatste truc gaat ongeveer als volgt: Na de eerste weigering barst de interviewer in snikken uit en vertelt dat dit al de zoveelste maal is dat hij geweigerd werd, dat hij per interview betaald wordt, het geld hard nodig heeft enzovoort. Vervolgens vraagt hij zich radeloos af wat er toch mis is. Heeft de respondent in spé de deur nu nog niet dichtgeslagen, dan is hij vrijwel verloren. De interviewer gaat nu namelijk de vraag stellen of de respondent hem wil helpen. Lukt dit, dan is de rest zeer eenvoudig. Hij krijgt eerst enige tips en stelt dan voor ze in een rollenspel te oefenen. De arme respondent kan zijn leerling nu niet laten vallen en stemt toe. Vervolgens zet de interviewer aanhalingstekens om rollenspel en neemt alsnog het interview af. In ons geval zal de schade 'slechts' psychisch zijn, het aantal encyclopedieën, horloges en vloerkleden dat echter op deze wijze verkocht is, zou van eenieder een vermogend man maken. Indien een respondent weigert omdat het onderwerp hem niet aanstaat, kan men verdere toelichting geven, bijvoorbeeld door te zeggen dat het ook heel belangrijk is dat tegenstanders aan het onderzoek meedoen.

Is de respondent bang voor de gevolgen van het onderzoek, dan kan

men hem verzekeren dat die er niet zijn (als dat waar is natuurlijk). Een laatste reden om te weigeren is dat de respondent geen tijd heeft. De enige manier die vrijwel altijd succes heeft om hier iets aan te doen is te proberen een afspraak te maken om een andere keer langs te komen.

Als de respondenten mee willen doen, moeten we, zoals bovengezgd, ook nog zorgen dat ze de *juiste antwoorden* geven. We hebben een aantal redenen genoemd waarom mensen niet mee willen doen, er zijn ook redenen waarom mensen juist graag mee willen doen. Deze redenen kunnen de juistheid van de antwoorden beïnvloeden. Een belangrijke reden is de eenzaamheid. Helaas is het zo dat in onze welvaartsmaatschappij de eenzaamheid een steeds groter probleem wordt. Duizenden, zo geen honderdduizenden vooral wat oudere mensen hebben zo weinig kans met iemand te praten, dat een interview gezien wordt als een prachtige kans even wat aanspraak te hebben. Het gevolg kan zijn dat de respondent het interview zoveel mogelijk probeert te rekken; ellenlange verhalen worden verteld over onderwerpen waarover men feitelijk geen mening of kennis bezit en veel meningen worden ter plekke bedacht. De interviewer dient hierop te letten en hierover een aantekening te maken, zodat men er in de verdere bewerking rekening mee kan houden. Ook de belofte om na het interview nog even te zullen blijven praten, kan positief werken. In de eerste plaats zal men het gesprek niet nodeloos rekken, dat kan immers van de tijd af gaan waarin, in ieder geval voor de geïnterviewde, interessantere onderwerpen aangeboord kunnen worden. En in de tweede plaats zal men zoveel mogelijk proberen mee te werken om de stemming erin te houden.

Deze soms overvloedige wens tot meewerken kan een ander 'lastig' verschijnsel veroorzaken, namelijk dat de respondent niet zozeer probeert juiste antwoorden te geven, als wel een zo goed mogelijke respondent te zijn. Meestal komt dit voort uit de wens de interviewer zoveel mogelijk te behagen, hetzij voor de interviewer zelf, hetzij om het onderzoek, althans in de ogen van de respondent, zo mooi mogelijk te maken. Dit verschijnsel werd aangetoond in een experiment waarin het ene deel van de respondenten met nauwelijks merkbare knikjes en nauwelijks hoorbaar goedkeurend gebrom en gehum werd 'beloond' voor een bepaald antwoord, terwijl een ander deel een dergelijke behandeling bij een geheel ander antwoord kreeg. De uitkomsten van beide groepen waren zeer verschillend. (Het effect van dit soort beloning werd ontdekt tijdens het



onderzoek naar de verbazingwekkende rekenkundige vermogens van het paard Hans. Pas na jarenlang onderzoek kwam een jonge assistent van een hoogleraar tot de ontdekking dat de trainer van het paard zeer kleine, nauwelijks zichtbare bewegingen maakte. Deze waren echter voor Hans voldoende om de hem gestelde problemen op te lossen en de wetenschappelijke wereld versteld te doen staan.) Een derde reden die onderzoekers doet geloven dat de antwoorden op de door hen gestelde vragen niet met de werkelijkheid overeenkomen, hangt samen met het doel van het onderzoek. Ter illustratie volgt hier zeer in het kort een onderzoek van Lapière uit 1935.

Lapière reisde een groot deel van Amerika rond in het gezelschap van een Chinees echtpaar. Zij bezochten ruim honderd hotels, restaurants, campings en dergelijke. In totaal werden ze slechts éénmaal geweigerd op grond van het ras van twee der reizigers. Later stuurde Lapière al deze gelegenheden een enquête (dit is een interview waarbij de respondent de vragen zelf leest en beantwoordt) waarin hij de vraag stelde of in het onderhavige etablissement Chinezen werden toegelaten. Slechts één respondente was bereid Chinezen toe te laten. (Zij refereerde overigens aan een bezoek van een leuk echtpaar, vermoedelijk de reisgezellen van Lapière.) Moeten we nu aannemen dat vrijwel alle respondenten hebben gelogen of zijn de resultaten van observatie en enquête beide waar en hebben we dus twee verschillende dingen gemeten?

In feite is dit de vraag naar ecologische validiteit, dat wil zeggen in hoeverre is een variabele, gemeten in een situatie waarin de respondent zich voor moet stellen dat hij iets doet, terug te vinden in het gedrag van die respondent in een natuurlijke omgeving. Een kant en klare oplossing voor dit probleem is niet te geven. Men zou de voorkeur kunnen geven aan observatie van de proefpersonen in de natuurlijke situatie, maar we kunnen moeilijk auto-ongelukken veroorzaken, als ons het gedrag tijdens auto-ongelukken interesseert. Een tussenoplossing bestaat uit het creëren van een situatie die lijkt op de toestand waarin men het te bestuderen gedrag normaal vertoont. Wil men bijvoorbeeld weten hoe iemand zich over een bepaald onderwerp zal uiten terwijl er anderen bij zijn, dan moet men deze persoon interviewen in aanwezigheid van anderen. Een voorbeeld van een onderzoek waarin deze aanpak werd gebruikt speelde zich af in een bar. Eén van de onderzoekers ging regelmatig naar deze bar. Zodra deze geaccepteerd werd, kwam een andere onderzoeker binnen en veroorzaakte problemen. De eerste onderzoeker kon op dat moment, zittend aan de tap, vrijwel ongemerkt

interviews afnemen over het gebeurde.

We moeten ons dus goed realiseren dat, als we iemand interviewen in zijn eigen huis, zonder toehoorders, terwijl de interviewer vertrouwenwekkend is en anonimiteit waarborgt, de respondent dan vaak uitspraken doet en gedrag voorspelt die niet overeen zullen komen met zijn werkelijke gedrag en uitspraken in situaties buiten zijn vertrouwde omgeving, met toehoorders. We zullen dit verschijnsel nogmaals tegenkomen bij de paragraaf over het formuleren van de vragen.

Ook allerlei kenmerken van de vragensteller kunnen van invloed zijn op de motivatie tot het geven van de juiste antwoorden. Geslacht, leeftijd, kleding, haardracht, ras, manier van spreken en doen en nog vele andere kenmerken kunnen de respondent doen besluiten de waarheid voor zich te houden. Dat kan heel onschuldig zijn, zoals het niet noemen van kleine dingetjes omdat de interviewer nogal gehaast overkomt, maar de antwoorden kunnen ook het tegenovergestelde van zijn mening zijn, zoals voor zou kunnen komen bij mannen die door een nogal potige dame geïnterviewd worden over feminisme. Ook hier geldt weer dat we ons eerst af moeten vragen wat we precies willen weten. Willen we bijvoorbeeld weten wat de blanke Nederlander denkt over Surinamers, dan kunnen we hem/haar het best fluisterend door een vertrouwenwekkende rasgeno(o)t(e) laten ondervragen. Willen we daarentegen weten of hij actief iets wil doen (meedoen aan een demonstratie van de Nederlandse Volksunie of zo) dan kunnen we hem beter laten interviewen door een Surinamer. (Zie bijvoorbeeld in Bovenkerk (1978) het onderzoek naar rasdiscriminatie op de arbeidsmarkt. Hierin liet hij afwisselend een Surinamer, een Spanjaard en een Nederlander de vraag stellen: 'Wilt u mij in dienst nemen'; zie hfdst. 9.) Het is dus niet zo dat we moeten proberen door allerlei maatregelen deze effecten te neutraliseren, we moeten er juist, indien mogelijk en nuttig, gebruik van maken.

Tot slot enige algemene aanwijzingen:

- a. Zorg voor goed getrainde interviewers.
- b. Geef ze goede richtlijnen mee.
- c. Leer ze zoveel mogelijk (volgens de onderzoeker ethisch verantwoord) methoden ter verhoging van de motivatie van de respondenten.
- d. Maak ze vertrouwd met de vragenlijst en indien mogelijk met het doel van het onderzoek. Niets is zo hinderlijk als een interviewer

die niet weet wat hij aan het doen is, of hoe hij verder moet. Soms is het ongewenst dat de interviewer het precieze doel van het onderzoek kent. Zo is er onderzoek gedaan naar de invloed van de eigen mening van de interviewer op de resultaten van een vraaggesprek. In zo'n geval kan men het echte doel niet noemen, maar dient men hem/haar een andere doelstelling op te geven, die niet in tegenspraak is met de vragenlijst zelf.

e. Zorg dat de antwoorden zoveel mogelijk letterlijk genoteerd worden en voorzien van alle gebeurtenissen die het antwoord kunnen verhelderen, zoals gebaren en geluiden gemaakt tijdens het interview. (Uiteraard gelden deze eisen ook voor de onderzoeker die zelf de interviews afneemt.)

### *Het formuleren van de vragen*

We hebben gezien dat men al veel problemen het hoofd moet bieden voordat de vragen gesteld kunnen worden. De vragen zelf moeten echter ook nog aan een aantal eisen voldoen.

1. In de eerste plaats moeten we bij het formuleren van de vragen letten op de kennis van de respondent, waarbij de woorden die gebruikt worden van groot belang zijn. De vragen moeten wat betreft woordkeus en zinsbouw zo gesteld worden dat de respondent ze begrijpt (ook al zou hij het zelf niet zo zeggen), en daarnaast moet uiteraard ook de interviewer de vraag begrijpen. Het aantal woorden dat de gemiddelde Nederlander kent is maar een klein deel van het totaal aantal Nederlandse woorden. Daarnaast zijn dat ook niet voor iedereen dezelfde woorden, zodat men in feite maar een beperkte keus heeft. Van die woorden die we gemeenschappelijk hebben, is er dan ook nog een aantal dat voor verschillende mensen verschillende betekenissen heeft; 'een mooi stuk land' betekent bijvoorbeeld voor een boer iets heel anders dan voor een toerist.

2. In de tweede plaats is het van belang dat de respondent de vraag binnen de goede context plaatst. Als we bijvoorbeeld een bekende de vraag: 'Hoe gaat het?' stellen, dan heeft deze vraag geen context. Als hij bijvoorbeeld zegt: 'Slecht' weten we eigenlijk alleen dat er iets slecht gaat, maar niet wat. Zijn antwoord kan slaan op zijn gezondheid of die van andere familieleden, op zijn werk, zijn auto of desnoods op zijn voetbalclub. Wordt de vraag daarentegen gesteld door een opzichter aan een arbeider, die een uur geleden een opdracht van hem heeft gekregen, dan zullen vraag en antwoord vrijwel zeker op die opdracht slaan.

Met dit probleem kunnen we drie dingen doen:

- a. we kunnen het gewoon vergeten;
- b. we kunnen proberen achter de context te komen die de respondent hanteert. De eenvoudigste oplossing is dan de vraag: 'Waarom?' te stellen. Zegt bijvoorbeeld een opvoeder dat hij zijn kinderen niet wenst te slaan, dan is het zeer belangrijk te weten waarom niet. Ziet hij het als een slecht pedagogisch middel, doet het hem meer pijn dan het kind of krijgt hij voortdurend klachten van de burens over geluidshinder door huilende kinderen;
- c. de derde mogelijkheid is de context zelf te controleren. We stellen dan niet de vraag: 'Hoe gaat het?' en vissen later uit, hoe de respondent de vraag heeft geïnterpreteerd, maar voegen aan de vraag een context toe: 'Hoe gaat het met uw gezondheid?'

De keuze welke van de oplossingen wordt toegepast, hangt vooral samen met het doel van het onderzoek. Zo is men bij verkiezingsonderzoek vaak helemaal niet geïnteresseerd in de 'waarom'-vraag. (Ook op vergaderingen tellen alle voor- en tegenstemmen even zwaar ongeacht of men stemt uit eigen belang, algemeen belang, of men door zijn stemgedrag iemand te vriend wil houden of juist wil dwarsbomen.) Ook de keuze tussen de tweede en derde oplossing zal men moeten maken op grond van het doel van het onderzoek. Bijvoorbeeld de volgende twee vragen behoren bij twee heel verschillende onderzoeksdoelen:

1. 'Hoe gaat het?' (Bij negatieve reactie volgt: 'Hoezo?')
2. 'Hoe gaat het met uw gezondheid?'

De eerste vraag kan passen in een onderzoek naar de vraag: 'Wat zijn de belangrijkste factoren die maken dat mensen zich niet goed voelen?' De tweede zou kunnen passen in een onderzoek naar de gezondheidstoestand van een bepaalde groep mensen.

Een derde fout die men kan maken, waardoor de waarde van de antwoorden kan dalen, is het gebruik van begrippen in de vragen, die de respondent wel kent, maar die hij niet goed kan hanteren. De begrippen introvert en extravert zijn over het algemeen wel bekend, maar het is voor vrijwel iedereen zeer moeilijk iemand introvert of extravert te noemen (tenzij de persoon in extreme vorm een van beide kenmerken bezit). Men kan deze moeilijkheid vermijden door de begrippen te vertalen naar een aantal dagelijkse gebeurtenissen: hoe gedraagt de betreffende persoon zich in groepen, op feestjes, als hij de weg kwijt is, in openbare gelegenheden enzovoort.

Een laatste eis die men aan de vragen stelt en die evenals de drie hierboven genoemde problemen tot doel heeft de vraag voor de

respondent en het antwoord voor de interviewer begrijpelijk en duidelijk te maken, is dat een vraag maar één idee mag bevatten. Bevat een vraag namelijk meerdere subvragen, dan weet men niet welke door de respondent beantwoord wordt. Men moet dan achteraf alle subvragen opnieuw en nu apart gaan stellen. Het is beter de vraag dan meteen te splitsen. (We zien dit verschijnsel ook vaak bij interviews aan politici. De interviewers stellen vaak vragen waarin vier, vijf of meer subvragen en vooronderstellingen zitten. De politicus zoekt dan meestal de vraag die voor hem het gunstigst is uit, praat daar langdurig over en weet zo de andere, lastige vragen te ontlopen.)

Ook het probleem of de respondent wel over de gevraagde kennis beschikt, kunnen we oplossen door het stellen van de juiste vragen. We hebben in de eerste paragraaf gezien dat er verschillende oorzaken kunnen zijn waardoor de respondent het antwoord op de vraag schuldig blijft, namelijk: hij heeft het nooit geweten (1) of hij is het vergeten (2).

1. In het eerste geval zijn er twee dingen die men kan doen:

a. Van te voren bedenken welke personen over de kennis beschikken. Dat betekent dat we bijvoorbeeld over het algemeen de heer des huizes zullen moeten ondervragen over de inkomende geldstroom en de vrouw over de uitgaande. Omdat dit echter niet in alle gezinnen op deze manier gaat, moeten we:

b. Het interview beginnen met de vraag of de betreffende persoon de beoogde kennis bezit. In dit geval dus: 'Bent u de kostwinner?' of 'Doet u het huishouden?' Soms is het niet nodig met zo'n vraag te beginnen. Het zou bijvoorbeeld geen beste indruk maken als u met de vraag: 'Eet u wel eens?' een interview over eetgewoonten begint. In andere gevallen zal het formuleren van zo'n vraag heel eenvoudig zijn, bijvoorbeeld: 'Heeft u bij de laatste verkiezingen gestemd?' of 'Kijkt u televisie?' Er zijn echter ook situaties waarin het heel moeilijk is zo'n beginvraag te stellen. Er zijn heel veel politieke en maatschappelijke onderwerpen waarover veel mensen nauwelijks een mening hebben, maar er zullen slechts weinigen zijn die dat toe willen geven. Eén van die dingen die men zeker niet moet doen, is het stellen van gesloten vragen (dat zijn vragen waarbij alle antwoordmogelijkheden gegeven worden). Maar ook open vragen geven geen 100% zekerheid. (We kennen allemaal wel iemand die elke willekeurige vraag beantwoordt met een stortvloed van woorden, ook als we van te voren weten dat de betreffende persoon volledig onkundig is op dat terrein.)

2. Tot slot nog enige tips om het geheugen van de respondenten op te frissen dan wel te controleren:

Om met het laatste te beginnen, er zijn heel veel gebeurtenissen die op de een of andere wijze vastliggen. Zo is het bouwjaar van een auto van het nummerbord af te lezen en zijn van belangrijke aankopen vaak nog gedateerde papieren in huis. Wil men dus exacte gegevens hebben en heeft men het idee dat er nog andere plaatsen zijn waar die te vinden zijn dan in het geheugen van de respondent, dan kan men daarnaar vragen.

Om het geheugen van de respondent *op te frissen* zijn meerdere manieren ontwikkeld:

a. Men kan de vraag een duidelijke context geven. Gaat het bijvoorbeeld om stemgedrag, dan kan men naast het jaar waarin de verkiezingen werden gehouden bijvoorbeeld de toenmalige lijsttrekkers in herinnering brengen, één of meer kwesties die toen van belang waren noemen of de uitslag geven. De respondent weet dan wat beter welke verkiezingen precies bedoeld worden en hopelijk is dan ook zijn antwoord dichterbij de waarheid.

b. Om het verschijnsel van de selectieve verschuivingen tegen te gaan (schulden en vliegpreizen, zie boven), kan men de respondent hierop wijzen in de vraag. Bijvoorbeeld: 'We weten wel dat voetbalprestaties in de herinnering steeds fraaier worden, maar we willen toch weten hoe u...'

c. Een methode die in sommige gevallen een gunstig effect heeft, is het stellen van vragen waarbij de respondent slechts iets hoeft te herkennen. Als we er bijvoorbeeld achter willen komen welke boeken over methoden en technieken docenten hebben gelezen, dan kunnen we ze een lijst geven waarop alle boeken staan die de laatste tien jaar zijn uitgekomen. Waarschijnlijk zullen zij geen boeken overslaan die ze wel gelezen hebben; het is wel mogelijk dat er een paar bijkomen. In andere gevallen zal het aanbieden van zo'n geheugensteuntje weinig uitkomst bieden. Wil men bijvoorbeeld weten op welke leeftijd een kind (dat nu volwassen is) zindelijk werd, dan is niet te verwachten dat een lijstje met alle leeftijden waarop een kind zindelijk kan worden zal helpen.

d. Een laatste mogelijkheid is het stellen van indirecte vragen. Deze methode past men toe als er van vergeten door verdringing sprake is. Men vraagt dan niet direct naar iets, maar probeert door het stellen van vragen naar dingen die ermee samenhangen toch het antwoord op de oorspronkelijke vraag te vinden.

Het laatste aspect waar men op moet letten is het verband tussen de

formulering van de vraag en de motivatie van de respondent. Sommige aspecten van de vraag kunnen een zeer slechte invloed op de motivatie van de respondent hebben. In de eerste plaats dient men denigrerende en schokkende woorden te vermijden. Sommige agrariërs wensen geen boer genoemd te worden (en andersom) en veel mensen vinden het niet prettig met vloeken en/of schuttingwoorden geconfronteerd te worden.

Ten tweede dient men over het algemeen geen vragen te stellen die vooronderstellen: 'Vindt u ook niet dat...'. De kans is groot dat de respondent onder invloed van de vraag instemt in plaats van zijn mening te geven. Het enige geval waarin men van dit verschijnsel gebruik kan en mag maken, doet zich voor als men een interview afneemt over onderwerpen waarbij sociale wenselijkheid een grote rol speelt, bijvoorbeeld: seksualiteit, criminaliteit en discriminatie. Alleen in dit soort gevallen kan het zijn dat men de vraag zo moet stellen dat men de respondent een bepaalde richting opdringt (dat moet dan wel tegen de sociale wenselijkheid in gebeuren). Zo'n vraag kan luiden: 'Wanneer heeft u voor het laatst in de gevangenis gezeten?' of 'Heeft u meer of minder dan tien maal overspel gepleegd?' Het antwoord op deze vragen zal waarschijnlijk dichter bij de waarheid liggen dan wanneer men eenvoudigweg had gevraagd of de respondent gezeten dan wel overspel gepleegd heeft. In de derde plaats dient men ervoor te zorgen dat de respondent het vraaggesprek niet naar zijn hand kan zetten. Om dit te verduidelijken het volgende voorbeeld: Eén van de auteurs moest ooit studenten interviewen over hun houding ten opzichte van de verschillende onderdelen van hun studie. Over elk onderdeel werd eerst gevraagd of ze tevreden waren. Zo ja, dan ging men over op het volgende onderdeel, zo nee, dan werd uitgebreid gevraagd wat er dan mis was en welke verbeteringen er zouden moeten komen. Eén van de respondenten ontdekte dit en op het moment dat het interview haar begon te vervelen, veranderde ze op slag van een ontevreden in een zeer tevreden studente. Het is dus aan te raden na beide antwoordmogelijkheden dóór te vragen, zodat de respondent niet naar believen het interview kan rekken of bekorten.

### *De vorm van de vragen*

We kunnen de volgende onderscheidingen maken:

1. Open – gesloten vragen
2. Directe – indirecte vragen

Daarnaast willen we in deze paragraaf nog ingaan op de volgorde van de vragen.

### *Open en gesloten vragen*

Gesloten vragen worden gekenmerkt door het feit dat in de vraag alle antwoordmogelijkheden besloten liggen. Veelal kunnen vragen op beide manieren gesteld worden, bijvoorbeeld:

Open: Hoe worden problemen binnen uw gezin opgelost?

Gesloten: Worden problemen binnen uw gezin meestal in onderling overleg opgelost, neemt één persoon de beslissing, of legt u het voor aan derden?

De discussie over de voor- en nadelen van open en gesloten vragen wordt al tientallen jaren met veel energie gevoerd. De laatste tijd zijn de meeste wetenschappers het er echter wel over eens dat deze discussie niet gevoerd kan worden zonder een aantal andere factoren erbij te betrekken. Vier zullen we er noemen:

#### 1. Het doel van het onderzoek.

Indien men niet beschikt over voldoende kennis betreffende het te onderzoeken gebied en men dus niet kan overzien of men alle antwoordmogelijkheden kent, dan dient men open vragen te stellen. Daarnaast is er het verschil of men diepgaand een klein gebied wenst te onderzoeken, of wat oppervlakkiger een groot gebied. In het tweede geval zal men over het algemeen veel vragen willen stellen. Om in dit geval de tijdsduur van het interview te beperken, kan men beter gesloten vragen stellen.

#### 2. De mate waarin de respondent over de gevraagde informatie beschikt.

Hierbij doet zich een paradoxale situatie voor, en wel deze: als de respondent over weinig informatie beschikt, dan kan men het best gesloten vragen stellen, zodat er niets verloren gaat. Tegelijkertijd bestaat echter het gevaar dat een respondent antwoorden geeft op vragen waar hij helemaal niets van weet. Men dient dus heel voorzichtig te zijn. Weet men zeker dat de respondent over veel informatie beschikt, dan komt dat er wel uit, ongeacht de vraagvorm. Is men er niet zeker van of de informatie aanwezig is, dan dient men zich er eerst goed van te overtuigen, dat die er wel is, voordat men gesloten vragen aanbiedt. Heeft men eenmaal een redelijke zekerheid, dan bieden gesloten vragen het voordeel dat ze de gedachten van de respondent beter structureren. Hij kent de mogelijke antwoorden en heeft dus meer houvast voor het kiezen van een antwoord.



### 3. Motivatie.

Ook hier zijn weer aan beide vormen voor- en nadelen verbonden. Het kost minder moeite gesloten vragen te beantwoorden en men hoeft zelden te zeggen dat men het antwoord niet weet. Daarentegen werkt het op sommige mensen frustrerend dat ze hun eigen mening niet kwijt kunnen. Veel mensen willen gedetailleerder antwoorden dan de aangeboden alternatieven toelaten.

### 4. De verwerking.

Tenslotte hebben open vragen voor de onderzoeker een nadeel. Voordat men kan gaan tellen en rekenen moeten eerst de antwoorden bewerkt worden (veelal met behulp van inhoudsanalyse). Antwoorden op gesloten vragen zijn automatisch al gecategoriseerd.

#### *Directe en indirecte vragen*

Het onderscheid direct-indirect slaat op het verschil tussen de te stellen vragen en het onderwerp van het onderzoek. Bij het reeds eerder aangehaalde voorbeeld over introversie-extraversie had de onderzoeker het idee dat directe vragen geen juiste antwoorden op zouden leveren. In plaats daarvan stelde hij indirecte vragen. In de vragen kwamen de woorden introvert en extravert niet voor. Het begrip is op verschillende manieren geoperationaliseerd. Hierin schuilt ook het grote gevaar van deze manier van vragen stellen. Er is namelijk over het algemeen zeer veel werk nodig om operationalisaties te controleren (*valideren*). Een ander voorbeeld is een vraag (sommigen noemen het zelfs een testje) om na te gaan hoe betrouwbaar iemand eruit ziet. Men legt daartoe iemand twee foto's voor en vraagt: 'Van wie zou u het liefst een tweedehands auto kopen?'

Wanneer gebruikt men nu indirecte vragen?

a. Zoals we al eerder zagen kan deze manier helpen om verdringing te omzeilen.

b. Indien men begrippen wenst te onderzoeken waarvan men het idee heeft dat de respondent deze niet of nauwelijks kent (dissonantie, identificatie, regressie enzovoort).

c. Indien het onderwerp zo emotioneel is dat de respondent het niet onder woorden wil of kan brengen.

d. Indien men materiaal wil verzamelen over respondenten, dat sociaal onaanvaardbaar is (luiheid, domheid, lafheid).

### *De volgorde van de vragen*

Over het algemeen is het het beste de onderwerpen zo te ordenen dat ze een voor de geïnterviewde logisch geheel vormen. Dit heeft het voordeel dat het interview meer het karakter van een gesprek krijgt en dat de respondent niet voortdurend van het ene op het andere onderwerp hoeft over te springen. (Een enkele keer, als men een spontaan antwoord wenst en geen overwogen oordeel, kan het nuttig zijn juist wel voortdurend van onderwerp te wisselen.)

Binnen de onderwerpen moet de volgorde van de vragen zodanig zijn dat het interview voor de respondent een logisch proces is. Vaak kan men de zogenaamde 'trecherteknik' toepassen. Dit houdt in dat men met de breedste (vaak open) vragen begint en daarna naar de meer specifieke, meer gedetailleerde gaat.

### *De voortest*

In de vorige paragrafen zijn tientallen problemen genoemd en richtlijnen gegeven voor het samenstellen van een vragenlijst. Bij het samenstellen van een vragenlijst zal de onderzoeker proberen zoveel mogelijk problemen te omzeilen en alle keuzen, die gemaakt moeten worden, te overwegen en te beargumenteren. De zo ontstane vragenlijst is echter nog steeds alleen getoetst aan het inzicht van de onderzoeker en de toepasbaarheid moet nog worden onderzocht. Daartoe is een voortest onontbeerlijk.

De voortest bestaat in ieder geval uit het afnemen van de vragenlijst. Daarnaast kan men na afloop een interview over het interview houden. Behalve de grote lijnen – of het geheel een beetje logisch overkwam, en de toon en de begrijpelijkheid van de vragen – kan men ook nagaan of merkwaardige of onwaarschijnlijke antwoorden veroorzaakt zijn door de formulering van de vragen of dat de geïnterviewde er een verklaring voor heeft. Bij een onderzoek naar het kopen van honde- en kattevoer bleek bijvoorbeeld dat één dame die niet in het bezit was van een kat, toch niet onaanzienlijke hoeveelheden blikjes kattevoer kocht. Bij navraag bleek dat ze vond dat een deel van haar burens hun katten verwaarloosde, zodat zij ze bijvoederde. Een onwaarschijnlijk maar juist antwoord. Een andere dame, in het bezit van één hond, verklaarde per dag zeven blikken hondevoer te kopen. Zij had de vraag niet goed begrepen, of die was niet goed gesteld. Ze bedoelde dat ze elke week op één dag zeven blikjes kocht.

Als respondenten kiest men mensen die zoveel mogelijk lijken op de later te onderzoeken groep. Vaak zal men tijdens dat vooronderzoek onduidelijkheden ontdekken die men achter het bureau over het hoofd heeft gezien. Vaak zal men vragen moeten veranderen, alternatieven toevoegen en soms zullen hele vragen moeten vervallen of worden toegevoegd. Indien men grote veranderingen heeft aangebracht, zal men opnieuw een voortest moeten uitvoeren. Pas als is gebleken dat alles in orde is, kan men met de echte afname beginnen.

### *Het houden van het vraaggesprek*

De rolverdeling bij het vraaggesprek is dat de interviewer het initiatief neemt en dat de respondent toestemt. In feite bewijst de respondent de interviewer dus een dienst. Het is belangrijk dit gedurende het vraaggesprek goed in het achterhoofd te houden. De onderzoeker kan nooit een gesprek eisen of antwoorden op bepaalde vragen, maar zal de situatie zo moeten maken dat de respondent wil meewerken. Dit begint al bij de introductie. De interviewer moet voldoende informatie geven om de respondent over te halen mee te doen.

Over het algemeen zal het noemen en, indien nodig, kort toelichten van de volgende punten voldoende zijn:

- a. Naam en functie van de interviewer.
- b. De instelling die het onderzoek uitvoert.
- c. Het onderwerp en het doel van het onderzoek.
- d. Waarom juist deze respondent gekozen is.
- e. Vertrouwelijkheid en anonimiteit van het gesprek (indien van toepassing natuurlijk).

Het is zaak deze introductie zo kort mogelijk te houden, zonder dat de respondent zich gedwongen of opgejaagd voelt en uiteraard ook zonder dat men de gehele vragenlijst ter goedkeuring behoeft voor te dragen. Eenmaal binnen zijn er drie dingen die men na moet streven. In de eerste plaats dat de respondent de juiste rol speelt. Hierover hebben we in het voorgaande al voldoende gezegd. In de tweede plaats dat de respondent zoveel mogelijk relevante informatie geeft en zo weinig mogelijk kletsboek. Dit alles met inachtneming van het derde punt: het handhaven van een sfeer die ervoor zorgt dat de respondent gemotiveerd blijft om met het interview door te gaan. In de loop van het gesprek kunnen een aantal technieken worden toegepast, die zonder de respondent te hinderen,

hem stimuleren in het beantwoorden van de vragen.

a. Het tonen van goedkeuring en interesse.

Over het algemeen zullen respondenten positief reageren op zinnestelsels als: goed zo, juist, dat begrijp ik, en goedkeurend gehum. We hebben echter ook gezien dat men de antwoorden op deze manier kan sturen. Het is dus van groot belang deze opmerkingen alleen op momenten te plaatsen dat dit niet kan gebeuren.

b. Het laten vallen van korte stiltes.

Als er gedurende een kort moment (2-3 seconden) een stilte valt, voelen de meeste respondenten dit als een uitnodiging verder te gaan, hun antwoord verder toe te lichten. Op deze manier heeft de gespreksleider niets te zeggen en zal hij dus ook de antwoorden niet beïnvloeden. Langere stiltes (10-15 seconden) werken vaak belemmerend.

c. Een andere manier om verdere toelichting op antwoorden te krijgen zonder dat men deze beïnvloedt, is het zogenaamde *spiegelen*.

Dit houdt in dat men het deel van het antwoord waarop men nog enige toelichting wenst, letterlijk herhaalt. Soms in de vraagvorm: 'U voelde zich toen niet zo gelukkig' of vergezeld van een vraag: 'U zei dat u zich toen niet zo gelukkig voelde. Kunt u dat toelichten?'

#### *Het vastleggen van de antwoorden*

Het vastleggen van de antwoorden is het laatste hier te behandelen maar zeker niet het onbelangrijkste onderwerp. Het eenvoudigst vast te leggen zijn de antwoorden op gesloten vragen. De antwoordcategorieën liggen van tevoren vast en meestal kan men dus volstaan met het zetten van een kruisje of een cirkeltje. Veel moeilijker is het vastleggen van de antwoorden op open vragen. We hebben reeds eerder gezien dat het van het grootste belang is dat de antwoorden letterlijk worden genoteerd. Alle methoden waarbij pas na afloop van het gesprek iets genoteerd wordt, zijn, indien te vermijden, uit den boze. Er blijven dan drie manieren over.

De eenvoudigste methode is het gebruiken van een bandopnameapparaat. Niets van het gesprokene gaat verloren (zolang men er tenminste voor zorgt dat de juiste knoppen ingedrukt worden en er voldoende band in het apparaat aanwezig is). Nadelen zijn dat veel mensen het bedreigend vinden (soms is dit door een gesprek vooraf wel te verhelpen) en dat het uitwerken zeer veel tijd kost (een uur

opname betekent ongeveer acht uur uitwerken).

De tweede methode is het letterlijk noteren van de antwoorden tijdens de afname. Het voordeel is dat veel mensen dit niet als bedreigend ervaren, maar daarentegen is het heel vervelend voortdurend tegen een schrijvend persoon aan te praten en vaak zullen er te lange pauzes vallen als de respondent uitgesproken maar de interviewer nog niet uitgeschreven is.

De derde methode heeft een nadeel: men krijgt *twee* nieuwsgierige mensen over de vloer, namelijk een interviewer en een notulist. De voordelen zijn dat men rustig kan spreken met iemand die je aankijkt en dat een notulist, die ook ingewerkt is binnen het onderzoek (vaak wisselen de notulist en interviewer per interview van rol), de interviewer kan aanvullen als deze iets vergeet.

### *Conclusie*

Zelfs een beknopte inleiding als deze maakt duidelijk dat het aantal fouten, dat men kan maken bij het gebruiken van het vraaggesprek als dataverzamelmethode, vrijwel niet te overzien is. Het verdient dan ook aanbeveling bij de keuze van een methode voor het verzamelen van gegevens niet blindelings naar de ogenschijnlijk eenvoudigste, het interview, te grijpen, maar dit slechts te doen op goede gronden en zo mogelijk de resultaten te controleren met behulp van andere methoden zoals inhoudsanalyse en observatie.

### *Aanbevolen literatuur*

Delen van de hieronder genoemde literatuur zullen u bekend voorkomen.

*Survey research center, Interviewer's manual.* Revised edition, Institute for social research, Michigan, 1976.

Ch. F. Cannell & R. L. Kahn, 'Interviewing'. In: G. Lindsey & E. Aronson, *The handbook of social psychology*, deel III. Reading, Mass., Addison-Wesley, 1969.

F. Bovenkerk (red.), *Omdat zij anders zijn.* Meppel, Boom, 1978.

#### 4. DE VRAGENLIJST

##### 1. *Inleiding*

Vragenlijsten worden zeer regelmatig gebruikt in sociaal-wetenschappelijk onderzoek. Behalve dat men ze door de respondenten zelf in kan laten vullen, worden ze ook vaak gebruikt als basis voor het gesloten interview. Een aantal criteria die van belang zijn bij het construeren van vragenlijsten, spelen ook een rol bij andere dataverzamelings technieken. Zo moet men zich bijvoorbeeld ook bij het open interview houden aan een aantal essentiële regels bij het formuleren van de vragen.

De beslissing om een vragenlijst als meetinstrument te hanteren, houdt in dat de onderzoeker voorbij is gegaan aan andere dataverzamelings technieken. Hij beschouwt, al dan niet op goede gronden, de vragenlijst als de beste keuze die hij ten behoeve van zijn onderzoek kan maken. Dat is nogal wat. De kans op vertekening en foutieve informatie is bij gebruik van de vragenlijst namelijk tamelijk groot. Het gaat immers om een rechtstreeks verzoek aan mensen iets te zeggen (in te vullen) over hun meningen, gevoelens, motivaties, zienswijzen, houdingen, waarden e.d. Onderzoekers die gebruik wensen te maken van de vragenlijst als onderzoeksinstrument, moeten dan ook heel voorzichtig zijn en zich steeds weer afvragen:

- a. In welke mate zou een vraag de respondenten (dat zijn de mensen die de vragenlijst invullen) ertoe kunnen brengen zichzelf in een gunstig daglicht te stellen?
- b. In welke mate zou een vraag de respondenten ertoe kunnen brengen rekening te houden met wat de onderzoeker (naar hun mening) graag zou willen horen?
- c. In welke mate zou een vraag de respondenten kunnen verleiden tot het geven van informatie waarover zij zelf niet of nauwelijks beschikken?

Natuurlijk gaan deze drie vragen over de betrouwbaarheid en validiteit van de vragenlijst als meetinstrument. Als onderzoeker behoort je aan de hand van deze vragen in eerste instantie na te gaan of voor de onderzoeksvraagstelling een andere methode misschien beter

was geweest. Als één of meer van deze vragen ontkennend of met een 'ik weet niet' moeten worden beantwoord, moet wellicht naar een andere benadering worden uitgezien. Bijvoorbeeld observatie, waar je niet bent aangewezen op *mededelingen* over gedrag, maar waar je dat gedrag rechtstreeks en onopvallend kunt waarnemen of registreren (vgl. hoofdstuk 2).

Hoe het ook zij, bepaalde soorten informatie kun je als onderzoeker alleen verwerven door ernaar te vragen. Het is voor een onderzoeker erg belangrijk te weten wat voor typen vragen er zijn en wanneer je ze moet gebruiken, en wat voor soort antwoord je dan kan verwachten. In dit hoofdstuk wordt hierop nader ingegaan.

Allereerst staan we echter stil bij de bouwstenen van het meetinstrument, de vragen zélf. Met name gaan we in op criteria voor het formuleren van (goede) vragen.

## 2. Criteria voor het formuleren van vragen

In de nu volgende paragraaf zullen we een aantal criteria behandelen die een rol spelen bij de constructie van een goede vragenlijst. (De indeling is van Segers (1977, hfdst. 6).) Uitgaande van het feit dat een vragenlijst erop gericht is de doelstellingen van het onderzoek te realiseren, zullen deze in de eerste plaats bepalend zijn voor de keuze van de vragen. Omdat de informatie echter van de respondent moet komen, moet ook aan diens kenmerken grote waarde worden gehecht. Dit criterium, *respondent orientation*, vereist een formulering die door de respondent (niet verkeerd) begrepen kan worden. Uit dit algemene criterium kan een aantal meer specifieke criteria worden afgeleid (Segers, o.c., blz. 153-161):

### A. *Het taalgebruik in de vragenlijst dient een volledige en nauwkeurige informatie te bevorderen.*

Een gedegen kennis van het spraakgebruik en de clichés van de respondent is nodig om gemotiveerd te kunnen beslissen in de vragenlijst bepaalde uitdrukkingen wel of niet te gebruiken. De taalkeuze kan verschillen al naargelang de te onderzoeken groep. Bij een grootscheeps onderzoek is de taalkeuze niet direct op een bepaalde groep te richten. Hier kan het beste worden gekozen voor eenvoudige formuleringen, waarbij oversimplificatie wordt vermeden. Het gaat niet alleen om adequaat taalgebruik, maar ook om de betekeniswereld die achter de woorden en uitdrukkingen schuilgaat. Een open vooronderzoek kan erg nuttig zijn voor het krijgen van inzicht in de betekeniswereld van de respondent.

B. *De vragenlijst moet ontworpen worden vanuit het referentiekader van de respondent.*

Vragen dienen zó geformuleerd te worden dat ze aansluiten op de leef- en denkwereld van de respondent. Inzicht in de referentiekaders van de respondent is een noodzakelijke voorwaarde voor adequaat geformuleerde vragen. Een voorbeeld kan dit verduidelijken. Huisvrouwen en studenten blijken de vraag: 'Hebt u in de afgelopen week werk in loondienst verricht?' dikwijls met 'Nee' te beantwoorden, ook als ze incidenteel of part-time wat geld hebben verdiend. De term 'loondienst' roept blijkbaar een ander referentiekader op dan de onderzoeker bedoelt, namelijk dat van een volledige dagtaak. Dit soort fouten is te voorkomen als je weet hoe de wereld van de respondenten eruit ziet en hoe daar gedacht wordt. Als men daar rekening mee houdt, kan dat bovendien motiverend werken, terwijl een vraag die niet klopt met de 'logica' van de respondent kan irriteren en de bereidheid tot antwoorden in het tegendeel doet verkeren.

C. *Een vraag mag niet uitgaan van irreële veronderstellingen over het informatieniveau van de respondent.*

Veronderstelt men meer kennis of deskundigheid bij de onderzochten dan in feite aanwezig is, dan kan dit makkelijk tot meetfouten leiden. De respondent zal ofwel proberen zijn onwetendheid te verbloemen door zo maar iets te antwoorden, of hij zal zich gefrustreerd voelen, hetgeen z'n bereidheid om te antwoorden doet afnemen. Vooral bij vragen die qua formulering of inhoud moeilijk zijn, kunnen deze problemen zich voordoen. Als een onderzoeker vermoedt dat een belangrijk aantal respondenten een bepaalde vraag niet kan beantwoorden, dient hij deze vraag zodanig te formuleren dat frustraties worden vermeden. Dit kan gebeuren door de vraag op een zodanige manier in te leiden dat de respondent duidelijk wordt dat lang niet iedereen het aangeroerde onderwerp kent. Hiermee doorbreekt de onderzoeker het verwachtingspatroon van de respondent: 'Wat aan mij gevraagd wordt, word ik geacht te weten.'

*Expert-error* is de vakterm voor meetfouten die optreden door het stellen van vragen die bij de onderzochte meer deskundigheid veronderstellen dan aanwezig is.

D. *Vragen moeten acceptabel zijn in het sociaal-culturele systeem waartoe de ondervraagde behoort.*

Vragen mogen geen bedreiging vormen voor de respondent. Denk aan onderwerpen uit de sociale taboe-sfeer. Tegenstrijdige gevoelens als: 'Hierover praat men niet' en 'Men behoort antwoord te



geven op gestelde vragen' plaatsen de respondent voor een dilemma. Het probleem speelt niet alleen bij de vraagstelling zelf, maar ook bij de reeks antwoordmogelijkheden die vaak bij de vraag wordt voorgelegd. 'Veilige' antwoordcategorieën zal hij misschien ten onrechte verkiezen boven sociaal 'ongepaste' alternatieven. Op die manier kan een flink vertekend beeld van een ondervraagde groep ontstaan. Een onderwerp dat moeilijk ligt, dient dus omzichtig benaderd te worden. Zoals onder (C) al werd aangegeven, kan ook hier een geruststellende, begrijpende inleiding effectief zijn. Onderwerpen die taboe zijn, wisselen in de loop van de tijd: vragen over seksualiteit roepen nu minder weerstand op dan tien jaar geleden. Bij dit type vragen dringt zich ook de vraag op naar de ethische toelaatbaarheid. In hoeverre mag de onderzoeker in de privé-wereld van de respondent binnendringen?

*E. De formulering van de vraag moet strikt neutraal zijn.*

Alle antwoordmogelijkheden dienen in principe een gelijke kans te hebben. De vragen mogen niet *suggestief* zijn, d.w.z. sommige antwoorden mogen niet beter of wenselijker lijken dan andere. Het noemen van één antwoordalternatief en het achterwege laten van andere is suggestief. Het genoemde alternatief verwijst in de ogen van de respondent naar verwachtingen van de onderzoeker, waaraan hij over het algemeen best tegemoet wil komen. Met het criterium van de neutraliteit is de metende functie van vragen in het geding. Een meetinstrument dat bepaalde waarden makkelijker registreert dan andere om redenen die in het instrument zelf gelegen zijn, is onbetrouwbaar en deugt niet. Het zondigen tegen het neutraliteitscriterium behoort tot de meest gemaakte fouten. 'U bent toch niet voor abortus?' is zeer suggestief geformuleerd. 'Bent u voor abortus?' is minder erg, maar helemaal goed is alleen: 'Bent u voor of tegen abortus?' Emotioneel en/of normatief geladen termen moeten omzichtig worden gehanteerd. Woorden met een scherpe 'lading' kunnen de eigenlijke bedoeling van een vraag maskeren, waardoor de antwoorden niet betrouwbaar zijn ('nazi', 'pedofiel' e.d.). Dit betekent niet dat emotioneel geladen zaken uit de vragenlijst moeten worden gebannen. Het betekent wel dat men rekening moet houden met de neveneffecten van het gebruik van dergelijke termen bij het bekijken van de gegeven antwoorden.

*F. Vragen moeten eenduidig en eendimensioneel zijn.*

Eendimensioneel betekent dat er slechts één ding tegelijk gevraagd wordt. Eenduidig wil zeggen dat de vraag door alle respondenten op dezelfde manier wordt begrepen. Als een vraag *meerduidig* is, weet

men niet meer wat de antwoorden betekenen. Een *meerdere* vraag leidt aan hetzelfde euvel. Immers, wat betekent het antwoord 'Ja, dat lijkt wel wat' op de vraag: 'Als het arbeidsbureau u voorstelt in een sociale werkplaats te gaan werken, zou u die kans aannemen of niet, of zou u iets voelen voor herscholing als u ondertussen voldoende inkomen zou krijgen?' Als de onderzoeker op elk onderwerp apart antwoord wil hebben, moet hij ook afzonderlijke vragen per onderwerp stellen.

Als in een onderzoeksvraagstelling meerdere eigenschappen of kenmerken van belang zijn, dient volgens criterium F dus per dimensie een afzonderlijke vraag gesteld te worden. Als de onderzoeker de relevante dimensies van een eigenschap nog niet kent, kan hij trachten deze op te sporen door er een zeer algemene, open vraag over te stellen.

G. *Bij de volgorde van de vragen moet rekening worden gehouden met psychische processen bij de respondent.*

Vragen moeten zodanig in volgorde worden gerangschikt, dat zij voor de respondent een zinnig geheel vormen en beantwoorden aan zijn gevoel voor 'logica' (vgl. criterium B). Dit kan erin resulteren dat vragen die in de theoretische gedachtegang (van de onderzoeker) dicht bij elkaar liggen, in een vragenlijst her en der verspreid zijn opgenomen. Voor het gevoel van de respondent moet het niet zo zijn dat hij van vraag tot vraag een gedachtensprong moet maken. De totale opbouw van de vragenlijst vereist een zorgvuldige aanpak. Afgezien daarvan kunnen er nog volgordekwesties zijn voor bepaalde onderdelen van de vragenlijst. Vragen over het bezitten van informatie of kennis van een bepaalde zaak zullen steeds vooraf moeten gaan aan vragen die opvattingen en meningen daarover beogen te registreren. Voor vragen die op een zelfde thema betrekking hebben, wordt de *funnel-approach* (trechtvormige vragenreeks) aanbevolen. Een voorbeeld:

- Wat vindt u van onze buitenlandse politiek?
- Wat vindt u van onze politiek inzake ontwikkelingslanden?
- Wat vindt u van onze politieke verhouding met Indonesië?

De gevolgde benaderingswijze houdt in dat meer algemene vragen worden gevolgd door meer specifieke. Dit om te voorkomen dat antwoorden op eerdere, meer specifieke vragen het antwoord op de volgende, meer algemene vraag beïnvloeden. Het beginnen met een algemene vraag laat meer ruimte voor 'ongedwongen' beantwoording van meer specifieke vragen op hetzelfde gebied dan bij de omgekeerde vraagvolgorde.

Een te abrupte overgang van het ene naar het andere onderwerp kan men ondervangen door een korte toelichting, waaruit blijkt dat men onderwerp A nu laat rusten en graag iets wil vragen over onderwerp B. Dit bevordert een soepel verloop van de vragenlijst.

Tot zover de criteria die de onderzoeker zou moeten hanteren bij de vraagformulering. Zoals in de inleiding werd opgemerkt, dient een onderzoeker ook te weten welke typen vragen er zijn en wanneer je ze moet gebruiken. Daaraan wordt in de volgende paragraaf aandacht besteed.

### 3. Soorten vragen

Een belangrijk onderscheid tussen diverse soorten vragen is de mate van *geslotenheid*. Een vraag is *open* als de respondent zelf het antwoord formuleert. De open vraag kan een belangrijke functie hebben in de voorbereidende fase van het onderzoek, wanneer het er om gaat zoveel mogelijk informatie in te winnen. Er zijn onderzoekers die vinden dat een open vraag in de *definitieve versie* van een schriftelijke enquête niet thuishoort (Bartelds e.a., 1978), omdat de kans groot is dat de daarmee ingewonnen informatie niet of nauwelijks te classificeren is. Vanuit een bepaald gezichtspunt, met name dat van de dataverwerking, is dit zeker te verdedigen. Toch kan het opnemen van enkele open vragen, ook in min of meer 'definitieve' vragenlijsten, een belangrijke rol vervullen. Niet alleen voor de onderzoeker, wanneer hij toch nog nieuwe aspecten aan een vraagstelling ontdekt die niet onmiddellijk in geprecodeerde antwoordcategorieën te vatten zijn, maar ook voor de respondent die op dit soort vragen spontaan kan reageren, hetgeen stimulerend kan werken op de beantwoording van de lijst als geheel. Toch komt de open vraag als dataverzamelmiddel beter tot zijn recht in een andere vraagmethode, namelijk het interview. Een goede interviewer kan aan de hand van een open vraag informatie verkrijgen die zeer genuanceerd is. Door op de juiste manier door te vragen kunnen voldoende specifieke antwoorden verkregen worden, die met een vragenlijst waarschijnlijk niet te achterhalen waren geweest. Interview en vragenlijst kunnen elkaar dus in bepaalde gevallen goed aanvullen.

In dit hoofdstuk zullen de voor- en nadelen van de open vraag nog meerdere malen aan bod komen. Voorlopig richten we de aandacht op het meest frequent gehanteerde type vraag (althans in de vragen-

lijst), de *gesloten* vraag, hetgeen betekent dat de respondent uit een aantal door de onderzoeker geformuleerde antwoorden moet kiezen. Binnen deze hoofdcategorie vragen zijn echter talloze onderverdelingen te maken (en ook reeds gemaakt). Het is onmogelijk ze binnen het bestek van dit hoofdstuk serieus aan bod te laten komen. We zullen ons dan ook beperken tot een aantal gangbare typen en nader ingaan op de overwegingen die een rol spelen bij de keuze van een bepaald type.

### 3.1. Veel voorkomende typen gesloten vragen

Bij gesloten vragen zijn de antwoordmogelijkheden beperkt en meestal ondergebracht in duidelijk onderscheiden categorieën (waarden van de variabele). Het gesloten karakter slaat eigenlijk meer op de antwoordmogelijkheden dan op de vraag zélf.

1. Vragen zonder door de onderzoeker aangeboden antwoordcategorie(ën), waarbij het antwoord echter wel 'vastligt' en dikwijls bestaat uit één woord of zinnetje (de Amerikaanse vakliteratuur spreekt treffend van *fill-in response*; in het Nederlands wordt wel gesproken van *schijnbaar open vragen*). Voorbeelden:

- Wat is het beroep van uw vader?
- Waar bent u geboren?
- Hoeveel kinderen hebt u?

De formulering van dit soort vragen maakt het zeer onwaarschijnlijk dat de respondent met een stortvloed van woorden of anderszins onverwacht zal reageren.

2. Vragen met twee of meer *feitelijke* antwoordmogelijkheden. Hier wordt bedoeld op antwoorden die een opsomming betreffen van kenmerken die op de respondent van toepassing kunnen zijn. Een paar voorbeelden:

- Ben je eerstejaars of ouderejaars student? (De antwoordalternatieven maken deel uit van de vraag.)
- Wat is uw netto-inkomen per maand?

Minder dan f 1200,-	1
f 1200,- tot f 1600,-	2
f 1600,- tot f 2000,-	3
f 2000,- tot f 2400,-	4
f 2400,- tot f 2800,-	5
meer dan f 2800,-	6

- Bent u gehuwd, ongehuwd of niet gehuwd maar wel gehuwd

geweest? (Ook hier maken de antwoorden deel uit van de vraag.)

gehuwd	1
ongehuwd	2
niet gehuwd maar wel gehuwd geweest	3

De respondent kan bij dit type vragen makkelijk geclassificeerd worden, hetgeen een snelle dataverwerking (computerverwerking) mogelijk maakt. Het is duidelijk dat in bovengenoemde voorbeelden de antwoordmogelijkheden elkaar uitsluiten. Dit hoeft niet altijd. Onder 5 zullen we een voorbeeld geven van antwoordcategorieën die elkaar niet uitsluiten. Eerst gaan we in op vragen die lijken op type 2, nu echter met evaluatieve antwoordmogelijkheden.

3. Vragen met twee of meer *evaluatieve* antwoordmogelijkheden. Het gaat hier om antwoorden die een *mening*, beoordeling of *standpunt* inhouden. De bijbehorende antwoordmogelijkheden zijn in aantal en in intensiteit variërende *categorieën over één onderwerp* (in tegenstelling tot de antwoordmogelijkheden bij type 2).

Voorbeelden:

- In welke mate bent u tevreden met uw huidige functie?

zeer tevreden	zeer ontevreden
------------------	--------------------

1	2	3	4	5
---	---	---	---	---

- In hoeverre bent u het eens met onderstaande uitspraak: 'Een man moet aan het hoofd van het gezin staan en de belangrijke beslissingen nemen'.

zeer mee eens	1
mee eens	2
geen mening	3
mee oneens	4
zeer mee oneens	5

- Bent u voor of tegen doodstraf?

ik ben voor	1
ik ben tegen	2

- Hoe is de sfeer onder de deelnemers aan het practicum doorgaans?

goed	1
goed noch slecht	2
slecht	3

Naar aanleiding van deze vier voorbeelden gaan we nader in op een

aantal belangrijke overwegingen die een rol spelen bij de formulering van de antwoordmogelijkheden van dit vraagtype.

Soms wordt een horizontale lijn gebruikt, waarbij alleen de uiterste waarden woordelijk zijn omschreven (zie voorbeeld 1, 'zeer tevreden' ... 'zeer ontevreden'), soms wordt ieder schaalpunt woordelijk omschreven en ontbreekt de lijn (vgl. voorbeeld 2 en 4). De antwoordmogelijkheden kunnen dus horizontaal en verticaal worden aangeboden en hun aantal kan variëren van twee (*dichotoom*) tot zeer veel. In beide gevallen spreekt men van een *schaal*. Hoeveel schaalpunten er zijn hangt af van de *complexiteit* van de vraagstelling. Bij eenvoudige vragen is bijvoorbeeld eens/oneens voldoende. Onder bepaalde voorwaarden kunnen de antwoorden op een aantal van dit soort vragen worden gecombineerd tot één waarde van een variabele. Over de verschillende manieren waarop men tot zo'n totaalscore voor een variabele kan komen, handelen de *schaaltechnieken*. Veel gebruikte 'schaaltypen' zijn de Likert-schaal en de Guttman-schaal (zie verder hoofdstuk 6). Een ander belangrijk punt is het al dan niet opnemen van een middencategorie (een 'neutraal' schaalpunt), dikwijls benoemd met 'geen mening', 'weet niet', 'maakt niet uit' e.d. Respondenten kunnen zo'n categorie als vluchtmogelijkheid beschouwen. De vraag in voorbeeld 2 over de positie van de man in het gezin had bijvoorbeeld ook 4 antwoordmogelijkheden kunnen bevatten ('geen mening' er uit), waardoor de respondent wordt gedwongen een standpunt in te nemen (vgl. de discussie in hoofdstuk 6 over de middenpositie). Bij de keuze van het aantal schaalpunten zou in algemene zin kunnen worden opgemerkt dat de onderzoeker zich dient af te vragen:

- in welke mate van gedetailleerdheid de respondent kan antwoorden en
- in welke mate gedetailleerdheid van de antwoorden gewenst is.

4. Vragen waarvan de antwoordmogelijkheden door de respondenten moeten worden *gerangordend*. Het toepassingsgebied van dit type vragen betreft waarderingen, voorkeuren, houdingen e.d. Soms wordt de respondent ook gevraagd zichzelf te plaatsen op een schaal.

Bij dit vraagtype krijgt elke antwoordmogelijkheid een positie toebedeeld in een rangorde van hoog naar laag. Het beantwoorden kost over het algemeen meer moeite en de vraag dient dan ook goed ingeleid te worden, eventueel voorzien van een meer of minder uitgebreide instructie.

Voorbeelden:

- Alle ouders willen dat hun kinderen bepaalde eigenschappen bezitten. Hieronder vindt u vier van deze eigenschappen. Kunt u de volgorde aangeven waarin u ze belangrijk vindt? (Denkt u aan kinderen van omstreeks 10 jaar.)

- a. Dat hij/zij rekening houdt met anderen.
- b. Dat hij/zij zelfbeheersing heeft.
- c. Dat hij/zij goed zijn/haar best doet op school.
- d. Dat hij/zij verstandig is en helder oordeelt.

Zoudt u in de kolom rechts van bovenstaande 4 eigenschappen een 1 willen zetten achter de volgens u belangrijkste eigenschap, een 2 achter de dan volgende, vervolgens een 3 en tenslotte een 4 achter de minst belangrijke?

Bij dit voorbeeld worden *alle* onderdelen van de vraag gerangschikt. Als er meer dan vier of vijf onderdelen zijn, kan het rangordenen leiden tot onbetrouwbare scores omdat de respondent te veel dingen tegelijk moet overzien. Als er bijvoorbeeld een oordeel wordt gevraagd over het belang van tien eigenschappen, kan de respondent gevraagd worden de belangrijkste er uit te kiezen. Je kunt ook vragen of hij de drie belangrijkste uitkiest en daarbinnen nog de volgorde van belangrijkheid aan wil geven. Dat verhoogt het informatiegehalte. De geringe betrouwbaarheid die bij de rangordening van veel categorieën tegelijk kan ontstaan, kan worden tegengegaan door de respondent niet *alle categorieën tegelijk* ter ordening voor te leggen, maar telkens slechts twee of drie (*partiële rangschikking*). Bij twee categorieën tegelijk spreekt men van *paarsgewijze vergelijking*, bij drie categorieën van *triatenvergelijking*. In hoofdstuk 6 wordt nader op deze methode ingegaan.

5. Vragen waarbij *meerdere antwoorden* kunnen worden gekozen (géén uitsluitende alternatieven/categorieën). De respondent wordt een aantal mogelijke antwoorden/alternatieven aangeboden, waarbij de vraag luidt *alle van toepassing* zijnde antwoorden aan te geven (men spreekt wel van een *checklist*).

Voorbeeld:

- Van welke vereniging(en) bent u lid? (Er kan meer dan één antwoord worden omcirkeld.)
- a. klaverjasclub
  - b. voetbalclub
  - c. hockeyclub

d. tennisclub

e. vormingscentrum

f. vakbond

g. hengelvereniging

h. geen van de hier genoemde

Tot zover een (zeker niet volledig) overzicht van gangbare gesloten vraagtypen. Over het algemeen is de gesloten vraag gemakkelijker te beantwoorden (dikwijls niet zo makkelijk te stellen), makkelijker te coderen en te analyseren, zeker met gebruikmaking van een effectief dataverwerkingsapparaat.

### 3.2. Verdere overwegingen bij het formuleren van vragen

Vaak wordt onderscheid gemaakt tussen *directe* en *indirecte* vragen. Het gaat er hierbij om in welke mate duidelijk is dat de vraag mikt op een specifieke informatie. In plaats van iemand te vragen hoe hij zijn werk vindt (direct), kan hem ook gevraagd worden of hij een werkloze vriend zou aanraden in het betreffende bedrijf te komen werken (indirect). De indirecte formulering beoogt openhartige, eerlijke antwoorden los te maken. Soms zijn er wat meer indirecte vragen nodig om de relevante informatie ten aanzien van een bepaalde zaak te verkrijgen. Als voor de vaststelling van iemands politieke overtuiging de indirecte methode wordt gekozen en op de vraag welke krant(en) en tijdschrift(en) worden gelezen geantwoord wordt *De Waarheid* en *Elseviers Weekblad*, zal men nog een paar vragen extra moeten stellen (voor meer informatie over (in)directe vragen, zie hoofdstuk 3).

Soms wordt bewust gebruik gemaakt van vragen die niet neutraal zijn geformuleerd, waardoor sommige antwoorden waarschijnlijker zijn dan andere. Vooral bij de constructie van schalen maakt men er nogal eens gebruik van. Het gaat hier om de zogenaamde *leading questions*. In plaats van te vragen: 'Houdt u wel of niet van uw werk?' vraagt men: 'Waarom houdt u van uw werk?' Het noemen van een autoriteit bevoordeelt bepaalde antwoorden ook boven andere. Soms wordt hier 'gebruik' van gemaakt: 'Het kabinet is van mening dat... Bent u het daarmee eens of niet?' Als men kritisch staat tegenover het kabinet, geeft men juist het tegenovergestelde antwoord. Hoe het ook zij, in beide gevallen treedt een vertekend beeld op.

Respondenten vertonen soms de neiging om volgens min of meer vaste patronen op vragen te reageren, ongeacht de inhoud van de vraag zelf. Ook bij niet-suggestieve vragen antwoorden nogal wat



mensen gemakkelijker met 'ja' dan met 'nee'. Deze neiging tot ja-zeggen is een voorbeeld van een *antwoordtendentie*. Een andere antwoordtendentie is *sociale wenselijkheid*, de neiging die antwoorden te geven die de respondent in een gunstig (of niet duidelijk ongunstig) daglicht stellen. Als de onderzoeker vermoedt dat bij bepaalde vragen dit soort tendenties zich kunnen voordoen, kan hij proberen gebruik te maken van andere antwoordcategorieën (bijvoorbeeld mee eens/niet mee eens in plaats van ja/nee) of andere formuleringen (bijvoorbeeld *indirecte*, ter vermindering van sociale wenselijkheid).

Tot slot nog een vergelijking tussen open en gesloten vragen. Aan het eind van de vorige paragraaf werd opgemerkt dat gesloten vragen in veel opzichten makkelijker zijn en het de onderzoeker mogelijk maken snel met zijn resultaten naar buiten te treden. Een probleem met open vragen is de geringere betrouwbaarheid. Het is min of meer 'toevallig' waar de respondent aan denkt bij de invulling van de vraag. Een maand later zou hij wellicht iets heel anders hebben ingevuld. De validiteit van open vragen kan soms gunstig afsteken tegen die van gesloten vragen, omdat de open vraag de respondent de gelegenheid geeft op een eigen directe wijze te reageren. Bij de gesloten vraag moet de respondent zich aansluiten bij antwoorden die door anderen voor hem bedacht zijn, hetgeen de validiteit negatief kan beïnvloeden. Open vragen vergen in ieder geval meer energie van de onderzoeker (bij de verwerking van de resultaten). Het stellen van drie of vier open vragen in een vragenlijstonderzoek bij 250 personen noodzaakt de onderzoeker eigenlijk tot het toepassen van inhoudsanalyse op de antwoorden. Daarom ziet men in 'definitieve' vragenlijsten bij grotere steekproeven weinig of geen open vragen. (Het stellen van gesloten vragen kost wel meer tijd dan open vragen, maar dat verschil is klein ten opzichte van het verschil in tijd dat nodig is voor de verwerking.)

#### 4. *Het grote belang van vooronderzoek*

Voordat de vragenlijst in het eigenlijke onderzoek wordt toegepast, behoort een gedegen vooronderzoek (*pilot study*) uitgevoerd te worden. De inhoud van de vragen in de vragenlijst is grotendeels gebaseerd op theorievorming en/of kennis van de resultaten van onderzoek op hetzelfde terrein. De relevantie van het vooronderzoek ligt dan ook voornamelijk op het vlak van de formulering en

vormgeving van de vragen en de vragenlijst als geheel. Vaak wordt een vooronderzoek uitgevoerd bij een groep respondenten die lijken op de uiteindelijk te onderzoeken groep (*populatie*), maar geen deel zullen uitmaken van de te trekken steekproef.

Een vooronderzoek bestaat uit diverse fasen. Zonder ze dwingend voor te willen schrijven, lijken de volgende stappen van belang:

- een reeks verkennende interviews die voorafgaan aan het formuleren van de vragen, mede bedoeld om zicht te krijgen op het referentiekader van de respondenten;
- constructie van een 'kladversie' van de vragenlijst, die getest wordt bij minstens 20 respondenten;
- de resultaten van de voorgaande stap dienen om de vragen aan te scherpen, bepaalde antwoorden te structureren en wellicht ook om een aantal vragen weg te laten. Tevens dient men vast te stellen of de vragenlijst het even goed doet in de diverse sociale klassen en/of culturele groepen binnen de totale te onderzoeken groep.

Binnen het kader van het vooronderzoek dient eigenlijk vastgesteld te worden of vragen open of gesloten moeten zijn, hoe de vragenlijst er uit moet zien, in welke volgorde de vragen gesteld moeten worden, hoe de afname of toezending het beste kan geschieden etc. Het kan niet genoeg herhaald worden: *vooronderzoek is essentieel!*

##### 5. *Een frequent toegepaste vorm van vragenlijstonderzoek: de postenquôte*

Bij een postenquôte wordt de vragenlijst toegestuurd aan de (potentiële) respondent. Deze moet de vragenlijst invullen en in een bijgevoegde antwoortenveloppe retourneren. De begeleidende brief is van groot belang. Er moet in staan wat het doel van het onderzoek is (vaak niet te gedetailleerd om beïnvloeding van de antwoorden te voorkomen), welke bescherming de respondent geniet (privacy, anonimiteit of niet e.d.), welke instantie/instituut het onderzoek uitvoert of ondersteunt, wie de onderzoeker is en wat zijn functie is. Verder moet de begeleidende brief een verzoek om medewerking bevatten (als het kan met uitleg waarom juist deze respondent(en) van belang is (zijn) voor het onderzoek) en zo nodig speciale instructies voor de invulling. Noem ook een uiterste datum van retournering en benadruk dat ze alle vragen beantwoorden.

Meestal is de postenquôte gestructureerd, weinig open vragen, omdat doorvragen niet mogelijk is. De mate van gestructureerdheid vereist vrij grote voorkennis van de onderzoeker over het onder-

werp. Bij de postenquête neemt het probleem van de *representativiteit* een centrale plaats in. Representativiteit kan worden omschreven als de mate waarin de onderzochte groep (de groep die de lijsten terugstuurt) dezelfde kenmerken bezit als de totale groep (*populatie*) waarnaar de resultaten worden gegeneraliseerd. Bedreiging van de representativiteit (*selectiviteit*) is bij dit onderzoek zo belangrijk, dat de voor- en nadelen van de enquête zelf er nauw mee verbonden zijn. Het is echter mogelijk door middel van respons-verhogende technieken het percentage respondenten dat reageert aanzienlijk te verhogen, soms tot 90%. Wat behoort er nu eigenlijk te gebeuren *nadat* de vragenlijst met de post is verzonden? Een en ander wordt kort besproken in het kader van de *follow-up*.

#### *Follow-up*

Na een periode van twee weken tot een maand wordt contact opgenomen met degenen die nog niet hebben gereageerd, de *non-respondenten*. Ook dit keer weer een brief met vragenlijst, inclusief antwoordenvolp (het origineel kan zoekgeraakt zijn). Gewoonlijk is tussen de 35 en 70% van de verzonden enquêtes binnen een maand geretourneerd. Zo'n 10 tot 25% kan worden gestimuleerd tot reageren door aan te dringen. Als de eerste follow-up brief nauwelijks effect heeft, sturen sommige onderzoekers een tweede, twee tot drie weken later. De meeste onderzoekers zijn niet bereid een respons van minder dan 75 tot 90% te accepteren (en terecht). Daarom stuurt men soms *nóg* een brief of neemt men telefonisch contact op. Waarom ook niet? Als een onderzoek de moeite waard is, dient naar de grootst mogelijke respons te worden gestreefd.

#### *Het trekken van een steekproef uit nonrespondenten*

Als minder dan 80% van de respondenten de vragenlijst ingevuld terugstuurt, moet een poging worden ondernomen gegevens te verkrijgen van een deel van degenen die niet reageerden (de zogenaamde nonrespondenten). Deze opzet slaagt als 5 tot 10% van de oorspronkelijke nonrespondenten alsnog de vragenlijst geheel of gedeeltelijk ingevuld terugstuurt. Het gaat er hierbij om vast te stellen of de nonrespondenten op de een of andere manier systematisch verschillen van de respondenten. Als deze controle niet wordt uitgevoerd, kan weinig gezegd worden over de interne of externe validiteit van de resultaten. Het uitvoeren van deze controle is niet eenvoudig. Mensen waarbij het tot twee of drie keer toe is mislukt ze bij het onderzoek in te schakelen, zijn moeilijk alsnog productief

te maken. Toch moet het geprobeerd worden. Eerst wordt 5 tot 10% van deze personen uit de lijst van nonrespondenten getrokken (op een aselechte, d.w.z. strikt toevallige, wijze). Vervolgens worden ze aangeschreven of opgebeld (soms zelfs bezocht). Omstreeks 75 à 80% van deze steekproef uit de nonrespondenten zal in het gunstigste geval reageren. Alles moet echter in het werk worden gesteld om dit te bereiken.

#### *Aanbevolen literatuur*

Delen van de hieronder genoemde literatuur zullen u bekend voorkomen.

Segers, J. H. G., *Sociologische Onderzoeksmethoden*. Assen/Amsterdam, van Gorcum, 1977 (hoofdstuk 6).

Bartelds, J. F., Kluiters, H. & K. G. van Smeden, *Enquête-adviesboek*. Groningen, Wolters-Noordhoff, 1978.

Tuckmann, B. W., *Conducting Educational Research*. New York, Harcourt Brace Jovanovich, 1978<sup>2</sup> (hoofdstuk 9).

Selltiz, C., Wrightsman, L. S. & Cook, S. W., *Research methods in social relations*. New York, Holt, Rinehart & Winston, 1976.

Warwick, D. P. & C. A. Lininger, *The sample survey: Theory and Practice*. New York, McGraw-Hill, 1975.

## 5. INHOUDSANALYSE

### *Inleiding*

Inhoudsanalyse is een techniek om iets over mensen en contacten tussen mensen te weten te komen, waarbij men als onderzoeksmateriaal in tegenstelling tot de andere methoden niet die mensen of contacten bestudeert, maar produkten van die mensen. Men neemt aan dat deze produkten bedoeld zijn om iets aan anderen mee te delen en daarom spreekt men meestal over boodschappen of communicaties.

Inhoudsanalyse is voor wetenschappen als geschiedkunde en letterkunde de belangrijkste dataverzamelmethode, vandaar dat er een aantal voorbeelden uit die gebieden is opgenomen. Het is natuurlijk niet zo, dat men, bedreven in de inhoudsanalyse, dan ook meteen letter- of geschiedkundig onderzoek kan doen. Elk onderzoek steunt op twee pijlers: de kennis van de techniek en de kennis van het onderwerp.

Veelal beperkt men zich tot geschreven boodschappen, maar vooral de laatste decennia is men deze techniek steeds meer gaan toepassen op andere vormen zoals gebaren, schilderijen, beeldhouwwerken, antieke Griekse vazen, films, melodieën en grafstenen. Ook komt het steeds vaker voor dat men materiaal gebruikt dat in het kader van het onderzoek is geproduceerd, zoals antwoorden op open interviews en psychotherapeutische gesprekken. Men zou dus kunnen zeggen dat elk menselijk produkt (cultuurprodukt) bruikbaar is. Aan inhoudsanalyse wordt veelal de eis gesteld dat ze objectief en systematisch is. Het eerste wil zeggen dat een andere onderzoeker, gebruik makend van hetzelfde materiaal en met dezelfde probleemstelling, tot dezelfde conclusies moet komen, het tweede dat alle materiaal ongeacht de inhoud gebruikt wordt, of, indien er een keuze gemaakt wordt, dat deze keuze niet afhangt van de vraag of de inhoud de onderzoeker bevalt of niet.

Daarnaast wordt er naar kwantificering gestreefd. Dit was tot voor kort een ondubbelzinnige eis, maar door veel onderzoekers wordt nu de waarde ingezien van kwalitatieve gegevens als aanvulling. Zoals Pool zegt: 'De opvatting dat kwalitatieve methoden inzicht

verschaffen en kwantitatieve alleen mechanische methoden voor het toetsen van hypothesen, is onjuist. De relatie is cirkelvormig: elk verschaft nieuwe inzichten die de ander tot nut kunnen zijn.<sup>1</sup>

### *Het doel van de inhoudsanalyse*

Het doel van deze techniek is het doen van uitspraken over de boodschap zelf of over andere variabelen die verband houden met die boodschap. De boodschap staat niet zo maar los, maar komt van een bepaalde persoon, die deze boodschap op een bepaalde manier tot iemand richt met een bepaald doel en met een mogelijk effect. We zien hier al een aantal variabelen waar we iets over zouden kunnen zeggen aan de hand van de boodschap.

### *De bron (zender, maker) van de boodschap*

Aan de boodschap is soms te zien wie de zender is. (Of zoals de Engelsen soms wat overdreven zeggen: 'All news are views'.) Men kan deze methode bijvoorbeeld gebruiken voor het bepalen van de schilder van een pas ontdekt ongesigneerd meesterwerk, of bij het beslechten van een ruzie tussen twee schrijvers die hetzelfde geschrift claimen. Aan de boodschap is soms ook te zien welke attitudes, waarden enz. de zender heeft of in welke psychische staat hij verkeerde. Een voorbeeld hiervan is de vergelijking van de inhoud van een artikel in een Amerikaans blad, dat verdacht werd van nazi-sympathieën, met propagandamateriaal van de nazi's zelf. Het feit dat zij qua inhoud sterk op elkaar leken, was reden voor de rechtbank de schrijver te veroordelen.

### *De vorm van de inhoud van de boodschap*

Veelvuldig is inhoudsanalyse toegepast om propagandamethoden te analyseren. Zo werd gekeken welke maatschappelijke waarden het meest voorkwamen in speeches van Kennedy, Nixon en Chroesjtsjov; tevens werd en wordt deze methode na vrijwel elke presidentscampagne in de Verenigde Staten toegepast. Ook stijlkenmerken zijn veel onderzocht. Men ontdekte bijvoorbeeld dat het aantal malen dat alliteratie in de sonnetten van Shakespeare voorkwam niet van kans afweek. Shakespeare bleek dus geen gebruik te hebben gemaakt van deze stijlfiguur. Een eenvoudig onderzoekje, dat de lezer zelf kan doen, is het verschil in stijl bekijken van een boek en een uittreksel. De hypothese kan bijvoorbeeld luiden: in een uittreksel komen relatief meer zelfstandige naamwoorden voor

dan in een boek. Veel onderzoek is, vooral in het verleden, verricht naar de leesbaarheid van teksten. Door het maken van een index gebaseerd op bijvoorbeeld woord en zinslengte en deze te normeren, is gepoogd op een eenvoudige manier te bepalen voor welke leeftijd of welk opleidingsniveau een bepaald boek, vaak een leerboek, geschikt was.

#### *Het onderwerp van de boodschap*

Hier kan men zeer veel kanten op. Men kan het gebruiken voor literatuurstudie als voorbereiding op een wetenschappelijk onderzoek; het beschrijven van de ontwikkeling van kennis op een bepaald gebied; om verschillen tussen landen wat betreft de inhoud van bijvoorbeeld kranteberichten over hetzelfde onderwerp te ontdekken; om verschillende media te vergelijken, bijvoorbeeld boeken en hun verfilming (het percentage overeenkomst varieerde tussen 38 en 89%). Voorbeelden van een aantal van deze mogelijkheden zijn te vinden in *The hermit of Peking* van Hugh Trevor Roper (1976). In dit boek probeert de schrijver het leven van Backhouse te reconstrueren aan de hand van allerlei documenten, zoals brieven, dagboeken, krante-artikelen en archiefstukken. Deze beschrijving vergeleek hij met de autobiografie van Backhouse (het onderwerp van boek en autobiografie). Daarnaast behandelt hij Backhouse ook als bron en probeert hij erachter te komen waarom deze zo'n merkwaardige en vrijwel geheel vervalste biografie schreef.

#### *De ontvanger*

Om uit de boodschap iets af te leiden met betrekking tot de ontvanger ervan moet men aannamen doen over hoe de boodschap tot stand is gebracht. De conclusie dat de normen en waarden, die behandeld worden in tijdschriften die veel door mensen uit respectievelijk hoge, midden en lage sociaal-economische milieus worden gekocht, sterk van elkaar verschillen, kan op meer dan één manier worden geïnterpreteerd: ofwel de schrijvers kennen de verschillende normen en waarden en schrijven zó dat de artikelen passen bij de beoogde lezers, of de lezers kopen de bladen als ze hun mening erin terugvinden, of de bladen veroorzaken die waarden en normen bij hun publiek. Dit laatste zou duiden op:

### *Het effect van de boodschap*

Het effect is niet zonder meer af te lezen aan de boodschap. Men moet meer informatie hebben. Met behulp van de resultaten uit het onderzoek naar de effecten van propagandatechnieken kan men voorspellen, tot op een bepaalde hoogte, welk effect een nieuwe boodschap zal hebben.

Meestal echter meet men het effect af aan de reacties van de ontvangers, bijvoorbeeld: ingezonden brieven aan een krant of telefonische reacties of opzeggingen na een televisie- of radiouitzending.

### *De maatschappelijke structuur*

Een laatste variabele (althans de laatste die hier genoemd wordt) is de maatschappelijke structuur waarin het geheel zich afspeelt. Een aardig voorbeeld is een onderzoek waarin gekeken werd naar de vragen en antwoorden in de vragenrubriek van een landelijk damesblad ('Margriet weet raad'). In de eerste plaats bleken de vragen vroeger veel meer over huishoudelijke probleempjes te gaan (hoe strijk ik de knieën uit de broek van mijn man enz.), terwijl op dit moment relatie- en huwelijksproblemen de meerderheid vormen. Hieruit bleek dat dit soort problemen nu veel beter bespreekbaar is. Ook de antwoorden verschilden sterk met betrekking tot het laatste soort problemen. Vroeger werd hoofdzakelijk geadviseerd het nog eens te proberen en werd de stereotiepe mannen- en vrouwenrol benadrukt; nu wordt steeds vaker een hulpverlenende instantie aangeraden, emancipatie gepropageerd en soms wordt, zij het vaak met enige terughoudendheid, het advies gegeven de problematische verhouding te verbreken. Hieruit zou men kunnen concluderen dat relatieproblemen in de huidige maatschappij opener besproken kunnen worden, dat het huwelijk niet meer zo onaantastbaar is en dat er wel degelijk in brede kring sprake is van enige emancipatie.

### *Wanneer past men inhoudsanalyse toe?*

Uiteraard is inhoudsanalyse de aangewezen techniek indien de boodschap zelf het onderwerp van onderzoek is. Als er bijvoorbeeld een twist is over wie iets heeft gemaakt, zal in ieder geval dat iets onderzocht moeten worden. Zo hebben vele deskundigen de schilderijen van Han van Meegeren onderzocht om te zien of het wel of niet Vermeers waren. (Overigens hebben ze naar verkeerde kenmerken gekeken: stijl, compositie, kleur enz. Volgens Van Meege-



ren had iedereen direct moeten zien dat zijn schilderijen aanzienlijk groter waren dan de tot dan bekende werken van Vermeer.) Ook bij de bepaling van leesbaarheid of het nakijken van opstellen en tentamens staan de boodschappen automatisch centraal.

In die gevallen, waarin men iets wil weten over andere variabelen (zender, ontvanger enz.), is het over het algemeen beter die variabelen ook direct te meten (interview, observatie). Soms is dat echter niet mogelijk of niet de beste manier, bijvoorbeeld als de persoon in kwestie overleden is. (In het reeds eerder genoemde geval van H. Trevor Roper was nog slechts één hoofdrolspeler in leven.) Hij moest zich dus hoofdzakelijk tot documenten bepalen. Het kan echter ook zijn dat inhoudsanalyse toegepast moet worden omdat de persoon niet bereikbaar is. Om wat meer inzicht te krijgen in de persoonlijkheid van A. Hitler, ten einde zo zijn toekomstige daden beter te kunnen voorspellen, werden zijn toespraken geanalyseerd; de geallieerden hadden weinig hoop dat hij zich aan een psychologisch onderzoek zou willen onderwerpen. Een tweede voorbeeld, ook uit de tweede wereldoorlog, is de analyse van onderschepte brieven van Duitse soldaten die in het belegerde Stalingrad zaten. Aan de hand van allerlei uitlatingen (bijvoorbeeld 'vaarwel' tegenover 'tot ziens') kon men de stemming schatten en ongeveer de te verwachten tegenstand bepalen. Een derde mogelijkheid is dat men verwacht dat de persoon geen juist verslag zal geven van het onderwerp. Zelfs sommige mensen die ooit lid waren van de NSB, om maar bij de tweede wereldoorlog te blijven, zeggen nu in het verzet te hebben gezeten. Door analyse van brieven uit die tijd, ledenbestanden enz. is hierover zekerheid te verkrijgen.

#### *Voor- en nadelen van inhoudsanalyse*

Zoals elke methode heeft inhoudsanalyse voor- en nadelen. Een aantal *nadelen* zijn:

a. Er moeten boodschappen zijn. Personen of volkeren die nooit boodschappen hebben nagelaten, zijn op deze manier niet te benaderen. Zo is de hele geschiedenis het verhaal van koningen, prinses, kardinalen, keizers enz. en niet dat van gewone mensen. Hoewel vooral de laatste tijd geprobeerd wordt hier enige verandering in te brengen, kleven hieraan allerlei problemen door de schaarste aan materiaal.

b. De validiteit van de methode is vaak nogal problematisch, vooral als men iets wil zeggen over andere variabelen dan die verbonden

aan de boodschap zelf. Met behulp van een batterij van zeven tests die, naar zeggen van de ontwerper, kon uitmaken of boeken door één of meer personen werden geschreven, kwam men tot de conclusie dat *Ulysses* door zes verschillende mensen werd geschreven en dat geen van allen de auteur was van *Portrait of the artist as a young man*. Hoewel James Joyce bewust verschillende stijlen gebruikte in *Ulysses*, moeten we toch concluderen dat de methode gefaald heeft, immers de ontwerper dacht een methode gevonden te hebben om schrijvers te identificeren, hoezeer ze ook hun best deden zich te 'vermommen'. In dit geval is hij hierin niet geslaagd (misschien is Joyce het enige genie dat dat kon, maar daarover bestaat geen zekerheid).

*Voordelen* heeft de inhoudsanalyse ook:

- a. Het zijn niets verstorende metingen, indien men werkt met bestaand materiaal. (In die gevallen, waarin men juist voor het onderzoek materiaal heeft laten produceren, geldt dit natuurlijk niet.) Het is bijvoorbeeld bekend dat de geschiedschrijving in de Sovjetunie nogal eens wordt aangepast aan de mening van de machtigste mensen van dat moment. (Trotzki bijvoorbeeld heeft nooit bestaan. In een museum in Moskou hangt een schilderij waarop hij naast Lenin staat afgebeeld. Vraag je de gidsen wie dat is, dan zeggen ze dat niemand dat weet.) Grafstenen echter bleken buiten deze 'correcties' te vallen. Een generaal die stierf in een periode waarin hij in ongenade gevallen was, heeft een heel klein eenvoudig grafsteentje. Zijn vrouw, die stierf toen de generaal weer gerehabiliteerd was, kreeg, omdat ze de vrouw van de generaal was, een imposant grafmonument.
- b. Het is een vrij goedkope methode. Elke bibliotheek beschikt over vele jaargangen van verschillende tijdschriften, kranten enz., die ieder ter beschikking staan.
- c. Gemaakte fouten zijn eenvoudig op te vangen. Het is eenvoudiger een nieuwe steekproef te nemen uit bijvoorbeeld de jaargangen van de *Libelle*, dan nogmaals een paar honderd mensen te gaan interviewen.
- d. Deze methode maakt de geschiedkunde en de letterkunde mogelijk. Hoewel er, zoals we hebben gezien, beperkingen zijn, is het de enige manier om vooral in het wat verder liggende verleden onderzoek te doen.

## *Hoe voert men een inhoudsanalyse uit?*

In het nu volgende deel zullen we systematisch de verschillende fasen van een onderzoek bespreken waarin inhoudsanalyse wordt toegepast. We zullen dit doen aan de hand van een onderzoek van Gottschalk en Gleser over de inhoud van zelfmoordbriefjes. Achtereenvolgens zullen we behandelen: het omschrijven van het probleem, het vaststellen van de eenheden, het maken van een categorieënschema, de voortest, het steekproeftrekken, de bewerking van het materiaal en het trekken van conclusies.

### *De omschrijving van het probleem*

Gottschalk en Gleser hebben zich jaren beziggehouden met het ontwerpen van een methode om persoonlijkheidskenmerken objectief meetbaar te maken door de analyse van gesproken en geschreven taal. Het hier beschreven onderzoek maakt deel uit van een groter onderzoek naar de persoonlijkheidskenmerken van mensen die zelfmoord hebben gepleegd. Door het verzamelen van kennis op dit gebied is het misschien mogelijk plannen tot zelfdoding eerder te herkennen en mogelijkere wijs te voorkomen.

Na het formuleren van de vraagstelling dient men zich af te vragen op welke manier men over het onderwerp informatie kan verzamelen. Het komt nogal eens voor dat onderzoekers elk probleem met dezelfde methode te lijf willen gaan, omdat die methode hen aanstaat. Ze zijn dan net als kleine kinderen die, nadat ze een verdoos hebben gekregen, tot de ontdekking komen dat alles een nieuw kleurtje nodig heeft. Men dient zich dus goed te realiseren dat het *probleem* centraal staat en niet de methode. Het te onderzoeken probleem vraagt om een methode. Kan en wil men meer methoden toepassen, dan is dat altijd voordelig, maar kies in de eerste plaats de best passende methode.

Bij een geval als zelfdoding zal het duidelijk zijn dat een groot aantal mogelijkheden uitgeschakeld is. Als de poging geslaagd is, zijn tests, observatie, vragenlijst en vraaggesprek met de betrokkene onmogelijk. Men kan zich dan alleen maar tot derden wenden, die iets over die betreffende persoon weten, of zoeken naar sporen (boodschappen) waarin men hoopt aanwijzingen te vinden over de geestesgesteldheid van de betrokkene op dat moment. De onderzoekers hebben gekozen voor de briefjes die een deel van de mensen schrijft vóór het plegen van de daad. Het spreekt vanzelf dat hier inhoudsanalyse de enig mogelijke keuze is.

### *De analyse-eenheid*

De analyse-eenheid is die eenheid waarover uiteindelijk een uitspraak wordt gedaan. In ons geval zijn dat die briefjes. Dat is een zeer beperkt gebied, vooral omdat maar in 15% van de gevallen een briefje achtergelaten wordt. Men kan zich veel grotere eenheden voorstellen, zoals alle schilderijen van Europese schilders vergeleken met die van Amerikanen, maar ook veel kleinere: zo zou men kunnen kijken of het woordje 'stuk' inderdaad een stuk mode is geworden.

Omdat het moeilijk is zo maar een uitspraak te doen zonder enig vergelijkingsmateriaal, hebben de onderzoekers de echte briefjes vergeleken met briefjes die geschreven werden door personen die in het geheel geen neiging tot zelfmoord vertoonden. (In hoeverre die mensen hieraan voldeden, zal de tijd moeten leren.) Hun werd verzocht een briefje te schrijven waarbij ze zich moesten voorstellen dat ze op het punt stonden dit leven vrijwillig te verlaten.

### *De tel-eenheid*

Dit is de kleinste eenheid die een bepaalde waarde krijgt. Soms valt de tel-eenheid samen met de analyse-eenheid, bijvoorbeeld als men de frequentie van bepaalde woorden telt en de resultaten gebruikt voor het construeren van een intelligentietest: de minst voorkomende woorden zijn het moeilijkst. Soms zijn ze zeer verschillend, men zou in een toespraak van de huidige minister-president (het is nu 1980) kunnen tellen hoe vaak hij de VVD en de PvdA noemt (onder het mom van: Waar het hart van vol is, loopt de mond van over). Hier zijn PvdA en VVD tel-eenheden, en is een toespraak de analyse-eenheid. Gottschalk en Gleser hebben woordjes geteld (hoe ze dat deden zullen we bij de categorieën verder bespreken). De tot nu toe genoemde voorbeelden zijn zeer eenvoudig en objectief te tellen, vaak neemt men echter minder duidelijke eenheden: bijvoorbeeld thema's, positieve en negatieve oordelen. Hierbij is het zaak deze eenheden zo duidelijk en begrijpelijk mogelijk te omschrijven, omdat anders aan de eis van objectiviteit niet meer wordt voldaan.

### *Meet-eenheid*

Met name in de jaren 1920 tot 1940 werd veel gebruik gemaakt van zeer objectieve meet-eenheden. Het belang dat verschillende kranten aan een bepaalde gebeurtenis hechtten, werd bijvoorbeeld bepaald door het aantal centimeters kolomlengte dat hieraan werd besteed. Onderwerpen in toespraken werden in seconden gemeten,

figuren op schilderijen in vierkante centimeters. Daarnaast telde men wel of een bepaald woord, thema en dergelijke voorkwam in een bepaald stuk. Men scoorde dan 'ja', ongeacht of het één of meer malen voorkwam (zie Observatie, tekensysteem). Soms werden ook frequenties geteld. Hoewel men steeds meer informatie verkreeg, bleven de resultaten toch erg mager. Om iets meer te kunnen zeggen, heeft men systemen ontworpen om ook iets over de richting en intensiteit van de tel-eenheden te kunnen zeggen. Eén van de best uitgewerkte systemen is dat van Osgood. Hij ontleedt elke zin in drie delen: het onderwerp waarover iets wordt gezegd, een kwalificatie van dat onderwerp en een deel dat beide verbindt, bestaande uit een werkwoord en in sommige gevallen één of meer bijwoorden. Beide laatste delen krijgen een positieve of negatieve en een zwakke of sterke waarde. (Bijvoorbeeld van +3 tot -3.) Het produkt van beide getallen geeft de score voor de zin. Deze kan in het gekozen voorbeeld lopen van -9 tot +9. De toekenning van de verschillende waarden aan de zinsdelen geeft veel problemen. We zullen deze behandelen in de volgende paragraaf, maar eerst een eenvoudig voorbeeld:

Eigengemaakte soep is lekker  
 $(+3) \times (+3) = 9$

Soep uit blik is soms lekker  
 $(-1) \times (+3) = -3$

Soep uit een pakje is bijna altijd vies  
 $(+2) \times (-3) = -6$

Omdat Gottschalk en Gleser geen idee hadden hoe ze de verschillende woordjes moesten waarderen (het onderzoek is opgezet om daar achter te komen), hebben zij voor een eenvoudige oplossing gekozen: de woordjes werden gewoon geteld.

### *Categorieën*

Zoals we al bij de tel-eenheid hebben gezegd, werden door de onderzoekers niet zo maar de woorden geteld, maar deze werden in bepaalde categorieën ondergebracht. Gottschalk en Gleser hadden daarvoor twee verschillende systemen ontworpen. Het ene was grammaticaal: werkwoorden, zelfstandige naamwoorden, tussenwerpsels enz., in totaal 7 groepen. De andere categorieën noemden zij psychologisch: gevoel of motivatie (angst, geluk, behoefte),

woorden die met waarnemen te maken hebben (kijken, voelen, jeuk), woorden die op nadenken wijzen (geloof, omdat, denk), en nog een groot aantal andere. Verder was er een aantal woorden die ze niet psychologisch konden duiden. Deze werden in een restcategorie gestopt.

Het grammaticale systeem is hetzelfde als dat wat we allemaal bij 'benoemen' op school hebben geleerd. Het tweede, psychologische, systeem moesten zij zelf ontwerpen.

Aan een goed categorie-systeem wordt een aantal eisen gesteld. Deze zijn:

a. De categorieën moeten uitputtend zijn. Aan elke eenheid moet men een categorie kunnen toewijzen. De eenvoudigste oplossing is het invoeren van een categorie 'overige' (zoals Gottschalk en Gleser ook gedaan hebben). Het gevaar bestaat echter dat, als de andere categorieën slechts een klein deel van het gebied bestrijken, 90% van de eenheden in de categorie 'overige' zal vallen. Omdat men dan nog niets weet, wordt als regel aangehouden dat de categorie 'overige' niet meer dan 5% van de scores mag bevatten.

b. Uitsluitendheid is een tweede eis. Het betekent dat elke eenheid slechts in één categorie mag passen. Men kan dit controleren door te kijken of goede codeurs dezelfde eenheid in dezelfde categorie stoppen. Klopt dit niet, dan zijn de categorieën niet wederzijds uitsluitend.

c. De verschillende categorieën moeten duidelijk geformuleerd zijn. De codeurs moeten niet eerst de categorieën hoeven te interpreteren voordat zij kunnen gaan scoren. Iedereen doet dat weer anders, zodat de resultaten niet meer betrouwbaar zijn.

d. Het systeem moet aangepast zijn aan het probleem. Als men alleen wil weten of men bepaalde zeepmiddelen kent, moet men geen systeem ontwerpen dat de diepgang van die kennis kan meten.

e. Het systeem moet bruikbaar zijn met betrekking tot het materiaal. Indien men weinig materiaal heeft, is het niet zinvol een systeem met een paar honderd categorieën te maken. Daarentegen zullen twee categorieën bij een grote hoeveelheid materiaal te weinig van de mogelijkheden uitbuiten.

Voor de constructie van een systeem kan men onder andere gebruik maken van schaalconstructiemethoden.

#### *De context-eenheid*

Dit is de grootste eenheid die men mag gebruiken als men moet

beslissen over de vraag welke waarde de tel-eenheid zal krijgen. Vooral als de tel-eenheid positief of negatief gescoord moet worden, zal het vaak nodig zijn een groter deel van het materiaal bij deze beslissing te betrekken. Is bijvoorbeeld een zin de tel-eenheid en wil men aan de hand van een artikel bekijken hoe iemand denkt over wijsgerige pedagogiek, dan zou men de zin: 'Wijsgerige pedagogiek is inhoudsloos geklets!' negatief scoren. Is het echter een aanhaling van een tegenstander, dan is dit zeker geen reden om aan te nemen dat de schrijver negatief staat tegenover het onderwerp.

Men kan als context-eenheid een alinea kiezen, maar ook een heel boek. Hoe groter de context-eenheid is, hoe moeilijker de beoordeling wordt. In ons voorbeeld zal de context-eenheid hoogstens een zin zijn, terwijl de meeste woorden zonder context in een categorie ondergebracht kunnen worden. Alleen woorden met een dubbele betekenis, zoals haar, pad, bleek, wel, waar, waard, maal, pot, want en roos kunnen zonder context niet geplaatst worden.

### *Voortest*

Voor bijna elke inhoudsanalyse moet een aangepast categorieënsysteem ontworpen worden. De problemen en het materiaal zijn zo verschillend dat het onmogelijk is een standaardstelsel te ontwerpen dat voor elk onderzoek geschikt is. Het is daarom zaak dit systeem en de mensen die ermee moeten werken vooraf goed te controleren. Men doet dit door het uitvoeren van een voortest. De voortest moet in alles vergelijkbaar zijn met de uiteindelijke uitvoering van het onderzoek. Drie dingen moeten na deze fase gecontroleerd en in orde zijn:

*a. De kwaliteit van de codeurs.* Vooral bij wat ingewikkelder systemen moeten ze aan twee eisen voldoen. In de eerste plaats moeten ze beschikken over voldoende taalvaardigheid. Hebben ze die niet, dan is de kans groot dat zeer veel eenheden in de categorie 'overige' of 'onbekend' terechtkomen. De tweede eis is een grote mate van accuratesse. Ze moeten voldoende geduld en vasthoudendheid hebben om telkens de juiste categorie te vinden en niet de eerste die erop lijkt te kiezen. De beste controle is het laten scoren van een standaardset. Men kan dan de ideale scoring vergelijken met hun prestatie.

b. *Het categorieënsysteem.* Zelfs al voldoet een systeem aan alle bovengestelde eisen, dan nog is het mogelijk dat er door de aard van het materiaal problemen rijzen. Het is gunstiger als deze in de voortest gesignaleerd en verbeterd worden, dan wanneer ze bij de uiteindelijke scoring te voorschijn komen. Al naar gelang het gevonden probleem kan het nodig zijn bestaande categorieën te preciseren, weg te laten, en/of nieuwe te creëren.

c. *De training van de codeur(s).* Zelfs goede codeurs, die met een goed systeem werken, maken fouten en als ze niet voldoende geoefend zijn, maken ze zelfs véél fouten. Als men door een aantal codeurs hetzelfde materiaal laat coderen, kan ofwel blijken dat ze op dezelfde categorie(ën) allemaal fouten maken, ofwel dat bepaalde codeurs meer fouten maken dan anderen en dat die fouten over alle categorieën verspreid zijn. In het eerste geval moeten die categorieën veranderd worden, in het tweede geval dienen de betreffende codeurs beter geoefend te worden, waarbij vooral een nabespreking tussen de codeurs onderling zeer veel kan verbeteren. De uiteindelijke betrouwbaarheid van de codering kan men met de methode van *intercoder agreement* bepalen\* (zie voor de berekening hoofdstuk 2).

Gottschalk en Gleser beschikten zowel over een goed getrainde codeur als over een goed getest systeem, zodat zij deze fase over konden slaan.

### *Het trekken van steekproeven*

Er zijn maar weinig onderwerpen waarover zo weinig materiaal voorhanden is, dat men al dat materiaal binnen één onderzoek kan analyseren. Meestal is er zoveel, dat men op de een of andere wijze een keuze dient te maken. Dit proces noemt men *steekproeftrekking* en het is tamelijk streng gereguleerd. Houdt men zich niet aan de regels, dan is de kans groot dat men een min of meer bijzonder deel van al het materiaal gaat bewerken. De conclusies kunnen dan niet generaliseerd worden. Analyseert men bijvoorbeeld alleen pagina 4 van *De Telegraaf*, dan zou de conclusie zijn dat het een tamelijk onschuldige kruising van een stripblad en een t.v.-gids is. In veel gevallen zal het steekproeftrekken stapsgewijs verlopen. In de eerste plaats moet men bepalen welk *medium* men kiest. Over sport kan men bijvoorbeeld iets te weten komen via kranten, radio, televisie, weekbladen, boeken en gedichten. Als evenwel in de



theorie al een medium genoemd wordt, dan kiest men dat medium. (Bijvoorbeeld: de *radionieuwsdienst* is de laatste tien jaren steeds meer aandacht gaan schenken aan verkeersongelukken.) In alle andere gevallen moet men óf alle media nemen óf een keuze doen; die keuze kan men dan het best door het lot laten bepalen. Deze loting kan gestratificeerd geschieden, d.w.z. indien men er zeker van wil zijn dat er evenveel leden uit de groep nieuwsvoorziening als uit de groep kunst zullen komen, dan deelt men de totale groep eerst in tweeën en trekt dan uit beide groepen evenveel leden. Binnen één medium zijn vaak weer verscheidene keuzemogelijkheden. Hetzelfde proces herhaalt zich dan. Als men kranten wil analyseren, kan men er bijvoorbeeld vijf trekken uit een hoed waarin alle kranten vertegenwoordigd zijn. Men kan ook hier weer bepaalde voorwaarden inbouwen. Elke krant kan bijvoorbeeld een gewogen kans krijgen al naar gelang zijn oplage. Heeft bijvoorbeeld *De Volkskrant* een oplage die tweemaal zo groot is als die van *Het Vrije Volk*, dan kan men *De Volkskrant* tweemaal zoveel lootjes in de hoed geven als *Het Vrije Volk*. Ook het maken van twee of meer groepen is mogelijk, bijvoorbeeld landelijke en regionale bladen. De volgende stap kan zijn het kiezen van een tijdsinterval. Dit kan één interval zijn, bijvoorbeeld mei en juni 1978; het kunnen ook meerdere intervallen zijn. Wil men trends in een bepaald blad bekijken tussen 1900 en 1975, dan zal men vaak een aantal intervallen kiezen, bijvoorbeeld 1900, 1910, 1920, ... 1970. Binnen zo'n tijdsinterval kan het materiaal nog te omvangrijk zijn. Men kiest dan bijvoorbeeld de krant van de 12de van elke maand. Bij deze keuze is het heel belangrijk op bepaalde cycli te letten. Een Amerikaans onderzoeker koos uit een groot aantal jaren de maand juni en analyseerde de trouwannonces. Hij kwam tot de ontdekking dat joodse Amerikanen niet trouwden. Bij nadere analyse bleek dat de maand juni vrijwel samenviel met een periode waarin het joden niet is toegestaan te trouwen.

De laatste stap is het kiezen, indien nodig, van het te analyseren deel van bijvoorbeeld de krant of het boek. Ook hier geldt weer dat men op de systematiek van het onderzochte moet letten. Neemt men bijvoorbeeld pagina 12 van elke krant, dan is de kans groot dat men uitsluitend advertenties vindt.

Concluderend kan men zeggen dat men bij elke stap eerst kijkt of er op grond van de theorie een reden is een besluit te nemen; indien dat niet het geval is, laat men de beslissing aan het lot over, waarbij men ervoor moet zorgen niet in cycli verstrikt te raken.

Ook Gottschalk en Gleser beschikten over meer materiaal dan zij wensten te analyseren. Zij moesten een keuze maken uit de in totaal 721 briefjes waar zij de hand op hadden kunnen leggen. Hun keuze viel op die briefjes waarvan de schrijver aan de volgende eisen moest voldoen: mannelijk, getrouwd, blank, protestant, geboren in Amerika, en tussen 25 en 59 jaar oud. Dat bleken er 33 te zijn. Vervolgens zijn zij 33 andere mannen gaan zoeken die ook aan deze eisen voldeden, en geen neiging tot zelfmoord vertoonden. Daarnaast werd ervoor gezorgd dat bij elke schrijver van een echt briefje iemand gevonden werd die qua leeftijd en beroep met hem overeenkwam.

Over deze manier van steekproeftrekken dienen we twee dingen op te merken:

a. Zoals al eerder gezegd, laat slechts een klein deel van de zelfmoordenaars een briefje achter (85% doet dat niet). De schrijvers van het artikel merken dan ook terecht op dat hun conclusies alleen gelden voor de onderzochte groep, totdat er bewezen is dat schrijvers en niet-schrijvers niet verschillen in persoonlijkheid.

b. Over de gevolgen van hun tweede stap (dat ze blank, getrouwd enz. moesten zijn) zeggen ze veel minder. Maar ook hier zullen we onze conclusies moeten beperken tot de groep die aan deze eisen voldoet, totdat bewezen is dat het ook voor anderen geldt.

Een andere vraag is *waarom* de auteurs zich deze beperking opgelegd hebben. Nu is dit niets unieks binnen de sociale wetenschappen. Deze truc wordt veelvuldig toegepast om ook (zeer) kleine verschillen zichtbaar te maken. Als een verschil tussen twee groepen groot is in vergelijking tot de verschillen binnen elk van de groepen, lijkt dat verschil heel belangrijk. Stel dat we iets willen zeggen over het verschil in lengte tussen mensen en we hebben dat alleen gemeten bij pygmeeën, dan komen we tot de conclusie dat mannen (de meeste althans) tussen 1,50 m en 1,60 m zijn, terwijl de vrouwen tussen 1,45 m en 1,55 m zijn. Een aanzienlijk verschil dus veroorzaakt door sekse-verschillen. Hebben we echter zowel bij de Masai als bij de pygmeeën gemeten, dan ziet het er anders uit. Mannen tussen 1,90 m en 1,50 m, vrouwen tussen 1,85 m en 1,45 m. Hoewel het verschil hetzelfde is gebleven, lijkt het veel minder belangrijk. Het toepassen van deze truc is dan ook heel gevaarlijk. Het lijkt dat we een groot deel van de verschillen kunnen verklaren, maar er kunnen andere variabelen zijn, die wij niet in het onderzoek opgenomen hebben, die veel belangrijker zijn. In het geval van de lengte leek eerst geslacht heel belangrijk (verklaarde

5 cm), maar het stammenverschil bleek veel groter (30 cm). We zullen aan de conclusie van Gottschalk en Gleser zien dat zij er, in ieder geval ten dele, rekening mee hebben gehouden.

### *De waardering van de gegevens*

De eerste stap na de codering van het materiaal, is een controle of alles wat beoogd werd verwerkt is. Hierna zal men proberen de grote hoeveelheid cijfertjes of turfjes te ordenen. Hiertoe staan alle statistische technieken de onderzoeker ter beschikking: rechte tellingen, tabellen, grafieken, gemiddelde, mediaan, correlaties enz. Eén waarschuwing is hier op zijn plaats. Hoewel de telling vaak op rationiveau geschiedt (bijvoorbeeld woordjes tellen), volgt daar zeker niet uit dat de variabele ook op dat niveau gemeten is. Als Van Agt in het genoemde voorbeeld 24 maal VVD en 3 maal PvdA zegt, wil dat niet zeggen dat hij 8 maal zo gek is op de VVD (aangenomen dat onze operationalisering klopt). We kunnen vermoedelijk niet verder gaan dan te zeggen dat hij meer voelt voor de VVD.

Gottschalk en Gleser berekenden per briefje het percentage woorden dat in elk van de categorieën viel. Vervolgens bekeken zij of er verschillen waren tussen die percentages voor echte en gesimuleerde briefjes. In totaal bleken tien categorieën een verschil op te leveren, vooral opmerkingen over mensen en dingen kwamen vaker voor in echte dan in namaakbriefjes.

Nu is het ook tijd enige controle op de operationalisatie uit te voeren, en wel door het bepalen van de validiteit. Dit kan op verschillende manieren. In de eerste plaats kunnen we kijken naar de *face-validiteit*, dat wil zeggen: heeft de operationalisatie zo op het oog iets met de te meten begrippen te maken. De onderzoeker zelf zal daar weinig moeite mee hebben, daarom kan men deze vraag beter voorleggen aan anderen. (Iedereen heeft ondertussen al gedacht dat het onderzoek over Van Agt niet klopte. Als hij de PvdA telkens roemt en de VVD vervloekt, komen wij tot een geheel andere conclusie dan zojuist.) In de tweede plaats kunnen we kijken naar de *concurrente validiteit*. Men bepaalt deze door een ander criterium te zoeken. Dit kan een andere boodschap zijn of gegevens verkregen met een andere methode van gegevens verzamelen. Om bijvoorbeeld uit te zoeken wie van twee schrijvers een bepaald stuk heeft geschreven, kunnen we naast inhoudsanalyse ook duim-schroeven gebruiken. De derde soort, *predictieve validiteit*, kan

slechts zelden gebruikt worden. Een van de weinige voorbeelden is het geval van de Noorse schrijver Hamsun. Op grond van zijn geschriften werden latente fascistische sympathieën ontdekt. Deze veronderstelling werd later bevestigd door zijn collaboratie met de nazi's.

Gottschalk en Gleser hebben de validiteit van hun methode getoetst door deze op nieuw materiaal toe te passen (*kruisvalidatie*). Deze steekproef week ook enigszins af van de eerder gebruikte, er waren nu namelijk ook vrouwen en alleenstaande mannen in opgenomen. Hoewel niet voor alle categorieën, bleek dat het grootste deel ook voor vrouwen en alleenstaande mannen hetzelfde verschil te zien gaf: het vaker noemen van mensen, dingen en plaatsen en het minder ter sprake brengen van gedachten en redeneringen. Dit geldt overigens alleen nog maar voor blanke protestanten die in Amerika geboren en tussen de 25 en 59 jaar zijn. Als men heeft geconstateerd dat betrouwbaarheid en validiteit voldoende hoog zijn, kan men overgaan tot het trekken van conclusies en de evaluatie van de theorie. Gottschalk en Gleser komen tot de volgende drie conclusies:

a. De theorie dat mensen, die op het punt staan zich het leven te benemen, overeenkomen met bepaalde psychiatrische patiënten (zoals een aantal theoretici had beweerd), die geen contact meer hebben met de wereld om hen heen, werd niet bevestigd. Het vaak noemen van mensen en dingen duidt eerder op een grotere betrokkenheid.

b. Het vaak noemen van plaatsen en dergelijke duidt volgens de onderzoekers op het besef dat deze mensen hebben dat ze een 'verre reis' gaan maken.

c. Het relatief ontbreken van gedachten en redeneringen duidt erop dat deze oplossing om van hun problemen af te komen meer emotioneel is dan beredeneerd.

In hoeverre deze conclusies algemeen geldig zijn t.a.v. ras, geboorteplaats, leeftijd en religie laten Gottschalk en Gleser in het midden; ze merken wel op dat ze alleen gelden voor mensen die een briefje achterlaten.

We hebben gezien dat bij gebruik van een goede theorie inhoudsanalyse meer kan opleveren dan de constatering (zonder aan de waarde van woordfrequentielijsten afbreuk te willen doen) dat onder 720.000 woorden, 'ik' 5037 maal voorkomt, waarvan 1645

maal geschreven met een hoofdletter. Daarentegen neemt de kans op fouten toe, zeker naarmate men van meer en onzekerder veronderstellingen uitgaat.

### Aanbevolen literatuur

Delen van de hieronder genoemde literatuur zullen u bekend voorkomen.

G. Wersig, *Inhaltsanalyse, Einführung in ihre Systematik und Literatur*. Berlijn, Verlag Volker Spiess, 1974.

O. R. Holsti, 'Content analysis', in: G. Lindsey & E. Aronson, *The handbook of social psychology*, deel III. Reading, Mass., Addison-Wesley Publishing Company, 1968.

H. Trevor Roper, *The hermit of Peking*. Harmondsworth, Penguin Books, 1976.

C. Brinkgreve & M. Korzec, *Margriet weet raad; gevoel, gedrag, moraal in Nederland, 1938-1978*. Utrecht, Het Spectrum, 1978.

L. A. Gottschalk & G. C. Gleser, 'An analysis of the verbal content of suicide notes', in: *British journal of medical psychology*, 1960, 33, 195-204.

### Noot

1. Geciteerd door Holsti (zie literatuur).

## 6. SCHAALCONSTRUCTIE

### *Inleiding*

In een groot aantal (vooral natuur-)wetenschappen is de relatie tussen een variabele en zijn meting zeer eenvoudig. In de economie kan men bijvoorbeeld inkomen operationaliseren als het aantal gulden dat iemand per maand verdient. De variabele en zijn meting zijn hier identiek. In één meting kunnen we de waarde van de variabele precies bepalen. Soms moeten we meerdere metingen doen om een variabele te meten. In de natuurkunde moet men hoogte, lengte en breedte vaststellen om de inhoud te bepalen. Desondanks is er geen verschil tussen de variabele en zijn meting. Binnen de sociale wetenschappen zijn er nauwelijks dit soort variabelen te vinden. De meeste variabelen zijn als de danseressen uit duizend-en-één-nacht en vertonen zich alleen verhuuld in vele sluiers. Om toch iets van die variabelen te weten te komen, zijn er methoden ontwikkeld, die het versluiserde beeld iets duidelijker kunnen maken. De belangrijkste reden voor de vaagheid van de variabelen binnen de sociale wetenschappen is dat ze min of meer geconstrueerd zijn en niet direct als eenheid voor ons verschijnen in de werkelijkheid. De mensen (meest wetenschappers) bemerkten dat bepaalde verschijnselen vaak samen optraden en plakten op het geheel een etiket. Soms is dat etiket een bestaand woord, soms een nieuw. Zo'n woord of *construct* duidt dan op een (mogelijk hypothetische) eigenschap van mensen die ze in meerdere of mindere mate kunnen bezitten. Die eigenschappen uiten zich op vele manieren in het gedrag. Een voorbeeld van zo'n *construct* is intelligentie. Niemand weet wat het is of waar het zit, maar iedereen weet ongeveer wat ermee bedoeld wordt en hoe het zich uit. Deze intelligentie wordt aan verschillende mensen in meerdere of mindere mate toegeschreven. De meting ervan gebeurt door niet-wetenschappers in veel verschillende situaties. Ze houden met van alles rekening om tot een oordeel te komen, zoals: opleiding, manier van spreken, manier van doen, beroep enz. En al die oordeeltjes worden gewogen en bewerkt tot ze een uiteindelijk oordeel geven: 'wel een slimme figuur'. Zo'n meting zal betrouwbaar, valide noch objectief

zijn. Vandaar dat er methoden zijn ontworpen om die meting systematisch te verrichten (zie ook hoofdstuk 7). De methode die wij hier behandelen is het zogenaamde *schalen*. Een schaal is in het kort een stel getallen (kwalificaties) die we op een voorgeschreven manier aan personen of stimuli toekennen, waarbij het bepalen van welk getal we toekennen aan een persoon of stimulus, afhankelijk is van de mate waarin die persoon of stimulus de eigenschap bezit die de schaal meet. Het toekennen van zo'n getal (*schaalwaarde*) aan een persoon of stimulus heet schalen. Voordat we aan de werkelijke constructie van schalen toekomen, moeten eerst enige onderscheidingen gemaakt worden die later van groot belang zijn.

a. In de eerste plaats moeten we bepalen wát we gaan schalen. We kunnen personen, stimuli of beide tegelijk schalen. In het geval dat we alleen *personen* willen schalen, nemen we aan dat alle stimuli (dat zijn vragen, plaatjes enz. waarop de proefpersonen moeten reageren) gelijk zijn. We kunnen bijvoorbeeld een twintigtal Leidse straten selecteren van ongeveer gelijke bekendheid (dat kunnen twintig hoofdstraten, maar ook twintig steegjes zijn). Hiermee zouden we de kennis kunnen meten die verschillende mensen hebben van Leiden. Hoe meer straten men kent, hoe beter de kennis van Leiden.

Schalen van *stimuli* wordt veel gedaan door marketing bureaus. Men vraagt een aantal huisvrouwen bijvoorbeeld welk blikje soep er het smakelijkst uitziet. We zijn dan niet geïnteresseerd in de verschillen tussen de vrouwen (we hopen zelfs dat die er nauwelijks zijn, in dit opzicht) maar in de vraag welk blikje het aantrekkelijkst is.

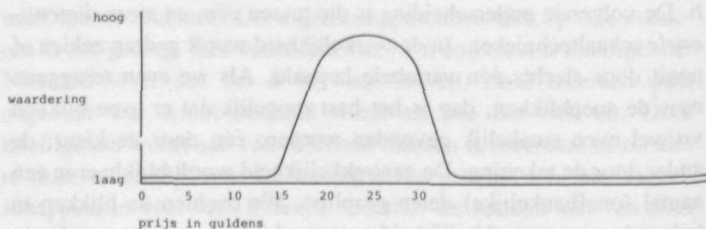
Het meest voorkomend is het geval waarin *zowel personen als stimuli* verschillen. Zo vinden verschillende mensen verschillende dranken lekker.

b. De volgende onderscheiding is die tussen *één- en meer-dimensionele* schaaltechnieken. In de werkelijkheid wordt gedrag zelden of nooit door slechts één variabele bepaald. Als we even teruggaan naar de soepblikken, dan is het best mogelijk dat er twee blikken vrijwel even smakelijk gevonden worden: één door de kleur, de ander door de tekening. De aantrekkelijkheid wordt blijkbaar in een aantal (onafhankelijke) delen gesplitst. We dachten de blikken in volgorde naar aantrekkelijkheid te rangschikken, maar ze werden in twee series gezet, die door elkaar lopen: een naar kleur en een naar tekening. Mogelijk dat ook andere dimensies meespeelden: tekst, vorm, grootte enz. Door de huisvrouwen deze verschillende dimen-

sies samen te laten vatten tot één oordeel, brengen we een bepaalde mate van willekeur in die de conclusie ondeugdelijk maakt. (We ontnemen onszelf ook de kans het ideale blik qua kleur, vorm, tekst enz. samen te stellen.) Het lijkt dus noodzakelijk de te meten variabele van te voren in een aantal dimensies te splitsen, zodanig dat we het hele gebied bestrijken. (Hiervoor bestaan allerlei data-analysmethoden zoals factor-analyse.) In het dagelijks leven doen we ook dergelijke dingen. Stel dat we bijvoorbeeld een boek willen kopen voor een jarige vriend. Enkele van de belangrijkste dimensies waar we dan rekening mee houden zijn: prijs, inhoud en levertijd (de volgorde is zuiver willekeurig). De prijs moet tussen bepaalde grenzen liggen, de inhoud moet aan een bepaald minimum voldoen en het boek moet geleverd worden voordat de vriend jarig is. Houden we na deze selectie nog meerdere boeken over, dan kunnen we extra dimensies invoeren: staan er plaatjes in, welk boek wil ik het liefst van hem lenen enz.? We zien dus dat we eerst moeten kijken hoeveel dimensies er zijn, dan de belangrijkste kiezen, daarna per dimensie een schaal maken.

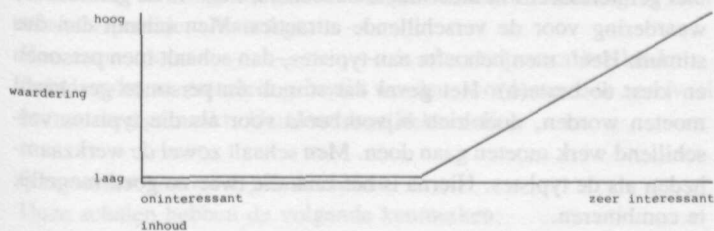
In dit hoofdstuk zullen we ons alleen bezighouden met één-dimensionele modellen. De meer-dimensionele modellen zijn in feite logische uitbreidingen en berusten op dezelfde principes.

c. Het derde onderscheid dat we maken betreft de karakteristieken van de variabelen. We willen aan elk niveau van de variabele een waarde toekennen. Om dit zinvol te doen moeten we iets weten over het verband tussen het niveau en de waarde die we toekennen. Bij de boekkoper van zojuist zagen we al enige voorbeelden. De prijs mocht niet te hoog en niet te laag zijn. De inhoud moest zo goed mogelijk zijn, maar in ieder geval aan een bepaald minimum voldoen. We kunnen dat grafisch als volgt weergeven:

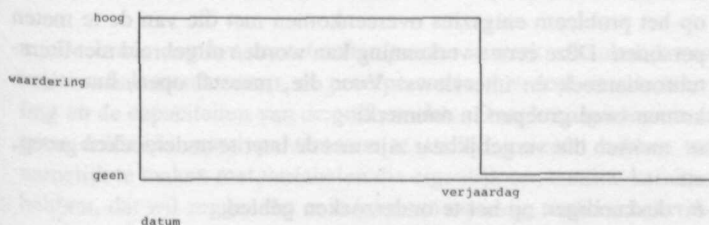


*Figuur 1.* Verband tussen waardering voor een boek en de prijs daarvan.



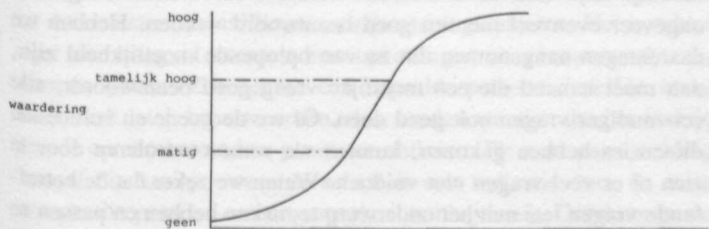


Figuur 2. Verband tussen inhoud en waardering van boeken.



Figuur 3. Verband tussen levertijd en waardering van boeken.

Dit zijn maar een aantal van de mogelijke vormen. Eén van de meest voorkomende is een soort platte S-vormige kromme.



De waardering gaat geleidelijk van geen naar matig, stijgt dan vrij snel naar tamelijk hoog en gaat dan heel langzaam naar hoog. Deze kromme vinden we op vele gebieden. Als we bijvoorbeeld prestatie verticaal en trainingstijd horizontaal zetten, krijgen we ongeveer dezelfde lijn. Om te bepalen welk model men zal gebruiken, is vaak langdurig vooronderzoek nodig.

De eerste onderscheiding: mensen, stimuli of beide, volgt meestal uit de vraagstelling. Bij het selecteren van circusattracties is men

niet geïnteresseerd in individuele bezoekers, maar in de gemiddelde waardering voor de verschillende attracties. Men schaaft dan dus stimuli. Heeft men behoefte aan typistes, dan schaaft men personen en kiest de beste(n). Het geval dat stimuli én personen geschaafd moeten worden, doet zich bijvoorbeeld voor als die typistes verschillend werk moeten gaan doen. Men schaaft zowel de werkzaamheden als de typistes. Hierna is het zaak die twee zo goed mogelijk te combineren.

De dimensionaliteit en de karakteristieken van de te meten variabelen zijn veel moeilijker te bepalen. In de eerste plaats zal men na moeten denken. Daar de onderzoeker ook een mens is, zal zijn kijk op het probleem enigszins overeenkomen met die van de te meten personen. Deze eerste verkenning kan worden uitgebreid met literatuuronderzoek en interviews. Voor die, meestal open, interviews komen twee groepen in aanmerking:

a. mensen die vergelijkbaar zijn met de later te onderzoeken groep, en

b. deskundigen op het te onderzoeken gebied.

Aan de hand van de zo vergaarde kennis kunnen we één of meer naar ons idee een-dimensionele schalen construeren. Door deze schalen aan een aantal mensen voor te leggen, kunnen we met behulp van statistische technieken controleren of onze ideeën kloppen. Als we bijvoorbeeld aangenomen hebben dat alle items even moeilijk zijn (zie de straten uit Leiden), moet elke vraag door ongeveer evenveel mensen goed beantwoord worden. Hebben we daarentegen aangenomen dat ze van oplopende moeilijkheid zijn, dan moet iemand die een moeilijke vraag goed beantwoordt, alle eenvoudiger vragen ook goed doen. Of we de goede en voldoende dimensies hebben gekozen, kunnen we soms controleren door te zien of er veel vragen niet voldoen. Weten we zeker dat de betreffende vragen iets met het onderwerp te maken hebben en passen ze toch niet in een reeds bestaande een-dimensionele schaal, dan vertegenwoordigen ze blijkbaar een andere dimensie, die we over het hoofd hebben gezien. We zullen nu wat dieper op verschillende schaaltypen en hun constructie ingaan.

### *Meer-puntsschalen*

Iedereen wordt op gezette tijden geconfronteerd met meer-puntsschalen. De bekendste zijn schoolcijfers (een 11-puntsschaal) en de bekende opinievragen:

Wilt u op de volgende schaal aangeven wat uw mening is over minister-president Van Agt.

Kan ermee door \_\_\_\_\_ is een ramp voor het land.

Behalve dat ze zeer veel gebruikt worden, worden er ook zeer veel fouten mee gemaakt. Vandaar deze behandeling.

### *Kenmerken van meer-puntsschalen*

Deze schalen hebben de volgende kenmerken:

- a. twee of meer schaalpunten;
- b. omschrijving van enkele of alle schaalpunten;
- c. meestal hebben ze een neutraal of middenpunt.

a. *De keuze van het aantal schaalpunten* is sterk afhankelijk van het te schalen onderwerp, de proefpersonen, het doel van de schaling en de capaciteiten van de onderzoeker. Over het algemeen is er geen duidelijk aantal punten vast te stellen, meestal hebben we namelijk te maken met variabelen die eigenlijk een continu karakter hebben, dat wil zeggen dat er in feite een oneindig aantal schakeringen zijn. Soms gebruikt men dan ook continue schalen, bijvoorbeeld:

Wat vindt u van geboorteverlof voor vaders?

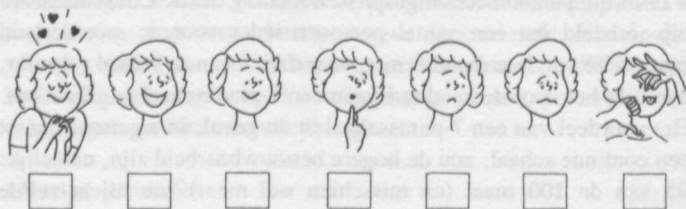
Zeer positief \_\_\_\_\_ zeer negatief.

Het is dan de bedoeling dat de respondent een kruisje zet op de plaats die met zijn mening overeenkomt. We kunnen dan, desnoods in duizendsten van millimeters, zijn mening nameten. Het is echter zeer de vraag of mensen hun opvattingen zo genuanceerd weer kunnen geven. Zouden we een bepaalde persoon dezelfde vraag heel vaak laten beantwoorden, dan zouden we zien dat het kruisje steeds op een iets andere plaats terechtkomt. De mate van spreiding van al die kruisjes over de schaal is een indicatie voor de mate waarin die persoon een uitgesproken mening heeft. Constateren we bijvoorbeeld dat een aantal personen ieder voor de meerderheid (zeg 95%) van hun reacties niet meer dan  $\frac{1}{7}$  van de schaal gebruikt, dan lijkt het voor de hand te liggen een 7-puntsschaal te gebruiken. Het voordeel van een 7-puntsschaal in dit geval, in tegenstelling tot een continue schaal, zou de hogere betrouwbaarheid zijn, namelijk: 95 van de 100 maal (en misschien wel meer) zou hij hetzelfde antwoord geven. Verder is ook de verwerking eenvoudiger. Uiteraard kunnen we iemand geen honderd of meer maal zo'n vraag laten beantwoorden. We kunnen echter ongeveer hetzelfde resultaat bereiken door honderd mensen de schaal tweemaal te laten aankruisen.

In het bovenstaande gingen we er vanuit dat we zo precies mogelijk wilden meten en stelden het aantal punten alleen afhankelijk van het onderwerp en de capaciteiten van de personen. Bij het toekennen van tentamencijfers zijn echter de docent en het doel belangrijker. Sommigen achten zich in staat een 41-puntsschaal te hanteren (6-, 7+ enz.), anderen 21 (6½, 9½) of 11 (4, 7). Veel scholen zijn overgegaan op een 4-puntsschaal, goed-voldoende-matig-slecht en slagen er toch op adequate wijze in slechte van goede presteerders te onderscheiden.

Een verschijnsel dat zeer vaak wordt geconstateerd, is het ontwijken van de schaaluiteinden. Er blijken altijd zeer weinig mensen te zijn die het ergens extreem mee eens of oneens zijn. Dit blijkt vaak niet zozeer met hun mening als wel met de procedure samen te hangen. Een remedie hiervoor is het invoeren van twee extra schaalpunten en deze zeer extreem te benoemen. Hoewel we dan een 9-puntsschaal aanbieden, meten we in feite toch op een 7-puntsschaal.

*b. Het omschrijven van de schaalpunten.* Het benoemen van de schaalpunten is bedoeld als steun voor de respondenten en kan gebruikt worden om het aantal punten uit te breiden zonder dat de betrouwbaarheid slechter wordt. De belangrijkste eis hierbij is dat men een duidelijk verschil moet zien tussen de punten. Het invoeren van bijvoorbeeld de punten: iets meer positief dan negatief, een beetje positief, zwak positief, eerder positief dan negatief, zal weinig helpen de meting preciezer te maken. De respondent zal nauwelijks of geen verschil zien tussen deze vier alternatieven. Naast omschrijvingen van de punten kunnen we ze ook illustreren met plaatjes, zoals in onderstaand voorbeeld. Het is heel moeilijk in woorden een even duidelijke omschrijving te geven.



Zet een kruisje in het vakje onder het plaatje dat het beste uitdrukt wat u vindt van het produkt dat u zojuist geproefd hebt.<sup>1</sup>



karakter van de poster te bepalen. Men zal eerst een tamelijk groot aantal posters selecteren (zeker meer dan men wenst te gebruiken in het experiment), waarvan men zelf vindt dat die posters het hele bereik van zeer aantrekkelijk tot angstaanjagend vertegenwoordigen. Deze worden dan aan anderen voorgelegd om ze te rangschikken. Voor de meeste (zo niet alle) mensen is deze opgave zeer moeilijk. Om dat op te vangen is de methode van de *paarsgewijze vergelijking* bedacht. Zoals de naam al zegt, worden de posters niet allemaal tegelijk aangeboden, maar in paren. We kiezen telkens uit de voorraad twee platen en laten de beoordelaars zeggen welke van de twee het meest bedreigend is. Het aantal mogelijke paren kunnen we berekenen met de formule

$${}_n C_r = \binom{n}{r} = \frac{n!}{(n-r)! r!}$$

( $n$  is het aantal posters,  $r$  is altijd twee in ons geval en  ${}_n C_r$  is het aantal mogelijke paren. ! achter een getal betekent dat men alle getallen in volgorde vanaf 1 t/m dat getal met elkaar moet vermenigvuldigen, dus:  $2! = 1 \times 2$ ;  $5! = 1 \times 2 \times 3 \times 4 \times 5$  en  $n! = 1 \times 2 \times 3 \dots \times (n-1) \times n$ .)

Het aantal combinaties loopt sterk op bij de toename van het aantal stimuli (bij 4 zijn het er

$$\frac{4!}{(4-2)!2!} = \frac{4!}{2! 2!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1 \times 2 \times 1} = \frac{24}{4} = 6,$$

bij 10 al 45 en bij 20, 190). Om een juist beeld te krijgen moeten alle combinaties vóórkomen en even vaak. Dit kan soms moeilijkheden opleveren. Als we bijvoorbeeld 50 stimuli willen ordenen, dan zouden de proefpersonen 1225 keuzen moeten maken. Dat is uiteraard onmogelijk. In ons experiment komen we al met een veel kleiner aantal in de problemen.

Het is, zeker bij zeer jonge peuters, niet goed mogelijk ze een direct oordeel te vragen. We moeten dus een andere methode ontwerpen. We zouden bijvoorbeeld kunnen kijken welke van de twee posters het kind langer bestudeert of naar welke het toegaat. In ons voorbeeld zou met tien posters het aantal metingen 45 worden. Dat zou lang gaan duren en zeer ongelukkige peuters en daardoor slechte gegevens opleveren. We kunnen dit probleem op twee manieren oplossen:

*a.* We schatten hoeveel keuzen we een proefpersoon kunnen laten maken en berekenen dan  $n$  uit de formule  ${}_n C_r$  ( ${}_n C_r$  en  $r$  zijn bekend, dus  $n$  is te berekenen).

b. In het begin hebben we de aanname gedaan dat alle proefpersonen onderling hetzelfde zijn. Dat betekent ook dat we bijvoorbeeld twee personen ieder de helft van het werk kunnen laten doen. Op die manier krijgen we toch voor elk paar een gelijk aantal metingen. Hoogstens moeten we, om een voldoende aantal metingen per paar te krijgen, het aantal proefpersonen uitbreiden.

#### De verwerking van de gegevens

Stel, om het verhaal niet al te ingewikkeld te maken, dat we de eerste oplossing hebben gekozen en 6 posters en 24 peuters hebben gebruikt.

We hebben dan  $\frac{6!}{4!2!} \times 24 = 15 \times 24 = 360$  waarnemingen.

De eerste stap is het plaatsen van deze waarnemingen in een tabel:

	1	2	3	4	5	6
1	12	4	2	6	10	0
2	20	12	8	11	16	4
3	22	16	12	15	21	2
4	18	13	9	12	11	6
5	14	8	3	13	12	7
6	24	20	22	18	17	12
	110	73	56	75	87	31

Langs beide assen worden de posters geplaatst (bijvoorbeeld 1 = John Travolta; 2 = Zwarte Piet enz.). In elke cel komt nu te staan hoe vaak de poster langs de horizontale as werd geprefereerd boven die langs de verticale as. In de omcirkelde cel staat dus dat John Travolta 20 maal werd geprefereerd boven Zwarte Piet. In de cel waar bijvoorbeeld John Travolta met zichzelf wordt vergeleken, een paar dat uiteraard niet wordt aangeboden, vullen we  $\frac{n}{2}$  in. We

nemen dus aan dat de proefpersonen John Travolta even vaak prefereren boven John Travolta als andersom. In ons geval komt daar dus  $24 : 2 = 12$  te staan. De volgende stap is het berekenen van de kolomtotalen. Deze getallen geven de volgorde aan van de posters betreffende hun dreiging. Het niveau van deze schaal is minimaal ordinaal. We weten zeker dat John Travolta het meest gewaardeerd wordt en dat nr. 6 (vermoedelijk Freek de Jonge) het meest dreigende gezicht heeft. De onbekend gebleven nummer 5 is

tweede. De derde plaats gaat met minimaal verschil aan de neus van Zwarte Piet voorbij en gaat naar nummer 4. Dit verschil is echter zeer klein en bij onderlinge vergelijking blijkt Piet zelfs iets beter uit de bus te komen (Piet wordt 13 van de 24 maal gekozen). Het lijkt het beste een keuze te maken tussen deze twee, tenslotte heeft het weinig zin stimuli aan te bieden met een even grote dreiging. In ons experiment kunnen we nu dus 5 verschillende stimuli aanbieden waarvan we de volgorde van bedreiging kennen.

In sommige onderzoeken kan het echter noodzakelijk zijn dat de stimuli een intervalschaal vormen. Om de gevonden schaal daartoe te verheffen is het nodig enige aannamen te doen. De geloofwaardigheid van de aannamen is sterk afhankelijk van het geschaalde kenmerk. Heeft men bijvoorbeeld op deze manier de lengte van verschillende objecten bepaald, dan zullen er weinig tegenwerpen zijn. In ons geval is dat echter al veel moeilijker waar te maken. In elk geval dient men niet al te lichtzinnig te zijn in deze beslissing. Een foutieve aanname kan een verder goed onderzoek toch rijp maken voor de prullenmand in plaats van publikatie. Voor de technische details van de verdere bewerking van de schaal tot intervalniveau zie men de boeken op de literatuurlijst.

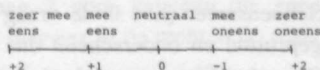
#### *De Likertschaal of de opgetelde oordelen*

Deze methode is min of meer het omgekeerde van de paarsgewijze vergelijking. Hij wordt toegepast als we personen willen rangordnen en ervan uitgaan dat de aangeboden stimuli een gelijke potentie bezitten om een bepaalde reactie op te wekken. Stel dat we een onderzoek willen doen naar de mate waarin een kind van zoetigheid houdt en het aantal gaatjes in zijn gebit. Het aantal gaatjes is eenvoudig na te gaan, zijn voorliefde voor zoetigheid echter niet. Eén van de methoden zou zijn te vragen hoeveel zoetigheid hij eet. Deze meting zal veel sociaal wenselijke antwoorden opleveren (er wordt veel stiekem gesnoept) en geen juist beeld geven. We kunnen dit probleem oplossen door een Likertschaal te construeren. We maken daartoe een lijst van een groot aantal zoete eetwaren: gebakjes, koekjes, drop, pepermint, jam, hagelslag enz. Over elk van deze eetwaren wordt een tamelijk extreme stelling geponeerd. Een deel in positieve zin bijvoorbeeld: drop is heerlijk, en een ander deel in negatieve zin: hagelslag is niet te eten. Elke stelling wordt dan voorzien van een meerpuntsschaal (vaak 5 of 7) waarop de respondent de mate waarin hij het eens is met de stelling, kan uit-

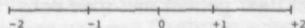


drukken. Door elk schaalpunt een waarde toe te kennen, kunnen we van elke proefpersoon per vraag een score vaststellen. De schaalwaarden lopen vaak van negatief naar positief, bijvoorbeeld:

Drop is heerlijk



Hagelslag is niet te eten



Men dient er goed op te letten dat aan vragen die negatief gesteld zijn, schaalwaarden toegekend worden in omgekeerde volgorde. De volgende stap is het optellen van alle scores per proefpersoon. Bij gebruik van 5-puntsschalen en twintig vragen kan de som-score dan lopen van  $-40$  tot  $+40$ . Nu dienen we na te gaan of alle vragen die we gesteld hebben, wel op dezelfde manier het kenmerk gemeten hebben. We selecteren daartoe, op grond van de som-scores, een groep die heel hoog en een groep die heel laag heeft gescoord (vaak neemt men de hoogste en laagste 25%). We nemen daarbij aan dat in ieder geval de meerderheid van de vragen datgene meet wat we willen meten. (Met deze methode controleren we dan ook niet de validiteit maar de *interne consistentie*, een vorm van betrouwbaarheid.) We hebben nu twee groepen, de ene van personen die veel van zoetigheid houden, de andere bestaande uit mensen die er niets van moeten hebben. Door voor beide groepen de gemiddelde score per vraag te berekenen, kunnen we zien of elk van de gekozen eetwaren door zoetekauwen hoger gewaardeerd wordt dan door hartigen. Het zal zelden voorkomen dat alle mensen uit de ene groep iets lekker vinden en niemand uit de andere groep. Meestal zal het verschil veel kleiner zijn, bijvoorbeeld 80% van de ene en 40% van de andere groep. Om te zien of dat verschil echt iets voorstelt of dat het ook door toeval ontstaan kan zijn, gebruiken we statistische toetsen. In sommige gevallen kunnen we zonder statistiek al iets zeggen. Bekijk bijvoorbeeld de volgende eetwaren en de bijbehorende appreciaties.

Gemiddelde mate van aantrekkelijkheid (-2 = vies, +2 = lekker).

	25% hoogst scorenden	25% laagst scorenden
schoensmeer	-2	-2
gekonfijte kersen	+1,9	-2
kaas	+1,2	+1,2

Schoensmeer (hoe is de vraag er trouwens ingekomen) wordt niet gepruimd en onderscheidt dus de twee groepen in het geheel niet. De gekonfijte kersen doen het juist prima. De zoetekauwen zijn vrijwel unaniem van mening dat die heerlijk zijn, de andere groep is het daar geheel niet mee eens. Kaas (weer zoiets vreemds) wordt in beide groepen even hoog gewaardeerd en voldoet dus ook niet aan ons doel. Door elke vraag door te lichten, houden we uiteindelijk een vragenlijst over, waarvan we weten dat alle vragen ongeveer hetzelfde meten. Na bepaling van de validiteit van de vragen kunnen we dan met enige hoop het onderzoek uit gaan voeren.

#### *De Guttman-schaal*

We bekijken hier alleen de Guttman-schaal om personen en stimuli tegelijk te schalen. De Guttman-schaal is niet de enige en ook niet altijd de beste methode, maar wel de eenvoudigste om een aantal kenmerken van dit type schaalconstructie te laten zien. Zoals altijd wanneer men een methode wil gebruiken die veel dingen kan, moet het materiaal waar die methode op gebruikt gaat worden aan veel en vaak strenge eisen voldoen. Voldoet het materiaal niet aan de eisen, dan moet men omzien naar een andere methode. In het geval van de Guttman-schaal zijn die eisen:

1. Op de te stellen vragen zijn maar twee antwoorden mogelijk.
2. De te stellen vragen meten hetzelfde kenmerk, maar in verschillende mate, dat wil zeggen dat sommige vragen door veel en andere vragen door weinig mensen bevestigend worden beantwoordt.
3. De te onderzoeken mensen moeten in verschillende mate over het te meten kenmerk beschikken.
4. Als men de vragen ordent naar moeilijkheidsgraad, dan moet er een vraag zijn voor iedere persoon, waarvoor geldt dat hij alle makkelijker vragen bevestigend en alle moeilijker vragen ontkenkend beantwoordt.

Een voorbeeld:

1. Heeft u de lagere school doorlopen?
2. Heeft u de middelbare school doorlopen?

3. Heeft u een universitaire titel?

4. Bent u gepromoveerd?

Het onderwerp en de bijbehorende vragen voldoen aan de vier eisen. Er zijn veel meer mensen die de lagere school doorlopen dan de middelbare enz. Verschillende mensen hebben een verschillend opleidingsniveau en voor iedereen is een vraag aan te wijzen, als bedoeld onder 4, bijvoorbeeld vraag 2 voor iedereen die geen ingenieur, doctorandus, dominee of meester in de rechten is, maar wel middelbare school heeft.

Dit voorbeeld is echter op twee punten weinig representatief voor de normale gang van zaken bij het maken van een Guttman-schaal.

1. Voor het meten van opleidingsniveaus is een eenvoudige vraag voldoende (welke opleiding heeft u?).

2. De schaal is zeer eenvoudig te maken (de vragen liggen nogal voor de hand).

Meestal zal het veel moeilijker zijn. Men kan de volgende problemen tegenkomen:

a. Meten alle vragen wel hetzelfde kenmerk?

b. Zijn de moeilijkheidsgraden (het percentage goede antwoorden per vraag) wel voldoende verschillend?

c. Is de volgorde qua moeilijkheid voor alle mensen hetzelfde?

Het eerste probleem zal door de onderzoeker zelf opgelost moeten worden. Er bestaan geen methoden en technieken om dat voor hem te doen. Hij zal dat dus op eigen inzicht, literatuurstudie en de mening van al of niet deskundigen moeten baseren. Eventueel kunnen bepaalde technieken te hulp geroepen worden om de gemaakte vragen enigszins te beoordelen. Nieuwe, betere vragen zal de onderzoeker echt helemaal zelf moeten verzinnen.

Voor de volgende twee problemen zijn wel technieken ontworpen. We kunnen dit illustreren aan de hand van een voorbeeld. Stel dat we willen bepalen hoe goed kinderen zijn in aardrijkskunde aan de hand van de kennis van hoofdsteden van verschillende landen door ze een lijstje met vragen voor te leggen als: Hoe heet de hoofdstad van... Deze opzet voldoet aan de gestelde voorwaarden. De vragen hebben inderdaad maar twee antwoordmogelijkheden, goed of fout. Ten tweede ligt het voor de hand dat hoofdsteden van kleine en verre landen minder bekend zijn dan die van de landen uit de Benelux. Verder zullen sommige kinderen er meer weten dan andere kinderen en zullen die zeker de bekendste kennen.

### *Dataverzameling*

Voor de dataverzameling maken we dus een lijst met vragen: Hoe heet de hoofdstad van... en leggen deze voor aan een groep kinderen die overeenkomt met de later te meten groep. Per kind wordt per vraag het antwoord gescoord: goed = 1, fout = 0. Vervolgens bepalen we per kind en per vraag het aantal goede antwoorden. Op grond hiervan kunnen we zowel de vragen ordenen naar hun moeilijkheid als de kinderen naar hun kennis. We kunnen nu alle gegevens in een zogenaamd *scalogram* zetten. Langs de verticale as staan de kinderen in volgorde naar het aantal goede antwoorden, bovenaan de laagst scorende en onderaan degene die er het meeste kende. Langs de horizontale as staan van links naar rechts de landen in volgorde van bekendste tot de meest onbekende.

Voorbeeld van een scalogram:

kinderen	Nederland	België	Amerika	Egypte	Jordanië	tot.
Jan	1	0	0	0	0	1
Miep	1	1	0	0	0	2
Klaas	1	0	1	0	0	2
Margreet	1	1	0	0	0	3
Koos	1	1	1	1	0	4
Jan II	1	1	1	0	1	4
Ria	1	1	0	1	1	4
Flip	1	1	1	1	1	5
tot.	8	6	5	3	3	

De moeilijkheidsgraden kunnen we aflezen aan de kolom totalen. Zo te zien behoeven we zeker geen vragen te verzinnen die nog eenvoudiger zijn, daarentegen zou het wel raadzaam zijn moeilijker vragen toe te voegen. (De moeilijkste vragen worden altijd nog door drie mensen goed beantwoord.) De aanpassing van de vragen aan het niveau van de leerlingen lijkt redelijk. De rij totalen lopen van 1 tot 5 (maximaal was 0-6 geweest). Tenslotte dient men te kijken of de moeilijkheidsgraad per leerling klopt. Daartoe kunnen we eerst het ideale patroon toekennen. Jan heeft één vraag goed, dat moet dan als het klopt de eerste zijn. We zetten dan een stippellijn (of een gekleurde lijn) achter Jan tussen Nederland en België. Zo gaan we verder. Achter Klaas (twee goed) zetten we dus een stippellijn tussen de tweede en de derde vraag. Als de vragenlijst ideaal is ingevuld, moeten nu onder de stippellijn alleen enen en boven die lijn alleen nullen staan. In ons voorbeeld blijken er drie nullen onder (en dus drie enen boven) de lijn te staan. Dat zijn dus

fouten in het patroon. Dat een schaal perfect is, verlangt niemand, de vraag echter hoeveel die schaal af mag wijken van het ideaal is veel moeilijker te beantwoorden. Het eenvoudigst is een bepaalde maat te ontwikkelen die die mate uitdrukt. We geven er hier een van, de zogenaamde representativiteit (R):

$$R = 1 - \frac{\text{aantal foute antwoorden}}{\text{totaal aantal antwoorden}}$$

(aantal foute antwoorden = aantal nullen boven + aantal enen onder de ideale lijn)

$$\text{In ons geval wordt } R = 1 - \frac{6}{40} = 0,85$$

Veelal vindt men dat een stel vragen en proefpersonen minstens een R van .90 moet hebben om nog als Guttman-schaal gebruikt te kunnen worden. Onze vragen en proefpersonen voldoen niet aan dat criterium. We kunnen nu twee dingen doen:

a. Vervelende proefpersonen weggooien. Klaas, Jan en Ria komen dan in aanmerking. Om dit te rechtvaardigen zullen we echter zeer goede gronden moeten hebben. Over het algemeen is een betere oplossing:

b. Vragen weggooien. In ons geval komen Amerika en Jordanië in aanmerking. Het is niet juist beide er meteen uit te gooien, omdat na verwijdering van één vraag de volgorde van de proefpersonen verandert. We verwijderen telkens het slechtste land totdat de R voldoet. Mocht het aantal vragen te klein worden, dan dienen we opnieuw te beginnen, met de goede oude vragen aangevuld met nieuwe. Het blijkt overigens dat Guttman-schalen vrij kort zijn.

### *Slot*

In dit hoofdstuk hebben wij ons beperkt tot slechts enkele eenvoudige methoden om een schaal te maken. In de literatuur vindt men, behalve nog veel meer soorten schaalconstructie, ook nog een groot aantal uitbreidingen (veelal sterk wiskundig) van de behandelde methoden. Indien men een schaal wil gaan construeren, is het dan ook onontbeerlijk de onderstaande boeken te raadplegen.

### Aanbevolen literatuur

Delen van de hieronder genoemde literatuur zullen u bekend voorkomen.

A. L. Edwards, *Techniques of attitude scale construction*. New York, Appleton-Century-Crofts Inc., 1957.

A. H. G. S. van der Ven, *Inleiding in de schaaltheorie*. Deventer, Van Loghum Slaterus, 1977.

### Noot

1. Tanur, J. M. e.a. (eds.), *Statistics: a guide to the unknown*. San Francisco, Holden-Day Inc., 1972, p. 225.

## 7. DE TEST

Lang geleden voelde een Chinese keizer zich van alle kanten bedreigd. Om zijn macht voor langere tijd te kunnen behouden, besloot hij een lijfwacht te vormen uit de beste krijgers die hij kon vinden. Helaas waren er in die tijd nog geen gespecialiseerde firma's voorhanden die gedrilde vechtmachines aan de hoogste bieder leverden. Iedere soldaat die benaderd werd, verklaarde – steeds even bescheiden – dat hij nu precies de man was die de keizer zocht.

Een vertrouweling van de keizer stelde zijne goddelijke majesteit dan ook voor zelf zijn mannen uit te kiezen door van redelijke afstand gevechten met de vijand te volgen en dan de meest dappere en kundige strijders, die nog in leven waren, uit te verkiezen voor zijn lijfwacht (*systematische observatie*). De keizer wees dit af; op deze wijze krijg je te weinig informatie om een gefundeerde keuze te doen, zei hij: de afstand waarop hij de strijd zou moeten volgen, zou immers veel te groot zijn. Een tweede idee van de vertrouweling, dat neerkwam op *participerende observatie*, kostte de man het hoofd (zie hoofdstuk 2).

De nieuwe vertrouweling van de keizer, een geleerde, ontwierp een fijn en degelijk plan. Hij vroeg de keizer een aantal eisen op te stellen, waaraan leden van zijn toekomstige lijfwacht dienden te voldoen.

De keizer kwam met de volgende eisen:

1. Met pijl en boog een man op vijftig passen kunnen doden.
2. In een man-tegen-man gevecht altijd winnen.
3. Te allen tijde bereid zijn voor de keizer te sterven.

Alle 11.000 opgeroepen sollicitanten verklaarden dat zij aan deze eisen zouden voldoen. Hoe het koren van het kaf te scheiden?

De geleerde raadsheer besloot de eisen zodanig op te stellen, dat ze getoetst konden worden (*operationaliseren*).

De eerste eis is al operationeel genoeg.

De tweede eis werd teruggebracht tot: In een man-tegen-man gevecht vier maal van een steeds sterkere tegenstander winnen. Immers 'altijd winnen' kan men niet testen, want dat kost een eeuwigheid en ontelbare slachtoffers.

De derde eis werd geoperationaliseerd tot: Het afleggen van de heilige gelofte om eerst het leven van de keizer te beschermen en daarna dat van zichzelf.

Nu bezat de raadsheer een *instrument* om de gestelde vraag – hoe selecteer ik een juiste lijfwacht – te beantwoorden.

De test werd afgenomen. De kandidaten kregen de opdracht op een afstand van vijftig passen op een voorbijganger te schieten. Dit kostte 1200 levens en 9800 teleurgestelde soldaten mochten weer naar hun respectievelijke leenheren vertrekken.

De 1200 overgebleven kandidaten werd verzocht zich een tegenstander uit te kiezen om een zwaardgevecht van man-tegen-man aan te vangen. De 600 overlevenden kregen in de tweede ronde dezelfde opdracht. Na een vorstelijke maaltijd volgden de derde en vierde ronde.

Alle 75 overgebleven moedige en kundige strijders legden tenslotte de heilige eed af om hun leven voor de keizer te geven.

Dat was het einde van een enigszins radicale *functie-geschiktheids-test*.

#### *De bruikbaarheid van een test*

Op een aantal punten kunnen we twijfelen aan de bruikbaarheid van deze test.

*In de eerste plaats* kunnen onder de 9800 afvallers bij het pijlschieten een aantal lieden geweest zijn, die, net toen ze de boog spanden, getroffen werden door een niesbui of op het moment suprême zodanig last van een vuiltje in het oog kregen, dat ze een minuscuul richtfoutje maakten, terwijl ze in feite evenzeer geschikt zouden zijn geweest als de uiteindelijke uitverkorenen.

Tijdelijke individuele gesteldheden kunnen onderdeel van de testsituatie uitmaken, waardoor niet het optimale testresultaat wordt behaald. Dit noemt men wel 'test-ruis'. Dit soort ruis kan vervelende gevolgen hebben: neem bijvoorbeeld een schoolleerling die een algemene niveautest moet afleggen, terwijl zij of hij juist een griepje onder de leden heeft of een emotionele gebeurtenis achter de rug.

Tests waarbij de geteste persoon (proefpersoon) erg zijn best moet doen (prestatie leveren) en die in een vrij korte tijdsperiode worden afgenomen, kunnen minder betrouwbare resultaten geven door deze ruis.

Enigszins kan deze ruis opgevangen worden door aan het eind van de testafname te informeren of de proefpersoon de test naar eigen



mening goed gemaakt heeft, of de proefpersoon zenuwachtig was, of iets dergelijks. Bij de testinterpretatie (advies, toelating, diagnose) kan men hier dan rekening mee houden.

Tests die weinig last hebben van ruis, noemt men *efficiënte* tests.

*In de tweede plaats* is het voor te stellen dat niet alle deelnemers onder gelijke omstandigheden getest zijn. De eerste boogschutters vuurden misschien hun pijlen af bij bladstil, helder weer, terwijl de laatsten hun pezen spanden in een ziedende tyfoon en met slechts 10 passen zicht. (Bij schaatswedstrijden doet zich dit sterk wisselen van de weersomstandigheden ook voor; iedereen weet hoezeer daardoor de resultaten beïnvloed worden.)

Iedereen moet onder dezelfde omstandigheden de test kunnen afleggen. Daarom moeten de testcondities nauwkeurig omschreven zijn om de vergelijkbaarheid tussen de verschillende testresultaten te garanderen. Pas als de vergelijkbaarheid aanwezig is, voldoet de test aan de eis van *standaardisatie*.

Let wel: Een test kan goed gestandaardiseerd zijn en toch inefficiënt.

*In de derde plaats* houdt de vergelijkbaarheid, naast de eis van standaardisatie, nog een andere voorwaarde in. Men moet een maat (norm) hebben waarop verschillen tussen te testen personen terug te vinden zijn. In de test van de raadsheer is de norm 'men voldoet aan de eisen of niet'. Een verder onderscheid tussen de deelnemers wordt niet van belang geacht.

De meest gebruikte tests kennen een meer genuanceerde normering. Deze normering is doorgaans tot stand gekomen op grond van frequente afname. Men verkrijgt aldus een schaal (denk aan de centimeterschaal), waarop geteste personen zich een positie verweren. Bijvoorbeeld: een rekencijfer  $8\frac{1}{2}$  impliceert dat de behaler van dit cijfer een betere prestatie geleverd heeft dan de persoon die een 7 behaalde. Normering houdt ook in dat vastgesteld is wat de gemiddelde testscore is (100 voor een intelligentietest) en hoe groot het percentage personen zal zijn tussen twee testcores. Voor de meest gebruikelijke intelligentietests geldt dat 67% van de mensen die de test afleggen, een score zullen behalen tussen de 85 en 115. Tests die niet met testcores in getallen werken, kunnen een normering hebben als: zeer zwak - zwak - gemiddeld - sterk - zeer sterk. Ook een normering naar categorieën komt voor: diagnose stellen bij ziekte.

Voor een algemeen gebruikte test geldt dat deze goed genormeerd is, maar bovenal dat aangegeven is *voor welke groep* (populatie) de test genormeerd is. Onderzoekers namen eens bij vele kinderen in de USA een test af, waarbij een mannetje getekend moest worden (Draw-a-man-test). De mate van expressieve intelligentie werd bepaald naarmate de kinderen een 'levensecht' tekeningetje van een mens kunnen maken. Met de indiaanse kindertjes bleek het zorgwekkend gesteld te zijn. Ze konden er niets van. De test was goed gestandaardiseerd en genormeerd (dacht men). Dus, hoewel de indiaanse kunstvoorwerpen tegenwoordig voor honderden dollars van de hand gaan, zijn indiaanse kinderen beduidend minder expressief intelligent dan de overige Amerikaanse kinderen. Toch konden de onderzoekers dat niet helemaal geloven en speurden naar oorzaken. Alras bleek dat indiaantjes uitstekend paarden konden tekenen, want dat waren ze van jongsaf aan zo gewend. De test was duidelijk genormeerd op niet-indiaanse Amerikaanse kinderen en kon dus niet geïnterpreteerd worden als indiaantjes de test aflegden. De test werd bijgesteld en heet nu: 'teken-een-mannetje-of-paard-test'.

Dus, een test dient goed genormeerd te zijn, dan weet je wat iemands relatieve score op de test is ten opzichte van de groep waarop de test genormeerd is.

*In de vierde plaats* is het legitiem zich af te vragen in hoeverre de Chinese keizer en/of zijn briljante raadsheer de uiteindelijke keuze van de 75 wachters heeft (hebben) beïnvloed. Wiste de keizer zich juist de tranen uit de ogen van het lachen om de hazelip van een der kampvechters, waardoor deze zich wat gegeneerd gevoelde en de degenhouw van zijn tegenstander niet kon pareren, of tikte de raadsheer even een der boogschutters op de arm, waardoor deze vragend opkeek, terwijl zijn pijl zich zovend ins Blaue een weg zocht?

Het spreekt, dat de onderzoeker die de test afneemt, op generlei wijze de geteste en daardoor het testresultaat mag beïnvloeden. In situaties waarin de onderzoeker individueel een test afneemt, mag zij of hij zich nimmer, hoe subtiel ook, laten leiden door gevoelens van sympathie, afkeer, medelijden etc. ten aanzien van de geteste. Hij of zij dient *objectief* te blijven. Ook bij de interpretatie van testresultaten, zeker waar het zogenaamde indirecte of projectietests betreft, bijv. Rorschach (inktvlekken) test, Thematic Apperception Test (op verschillende manieren te interpreteren plaatjes), Drie

Bomen-test, constructietests, associatietests, preferentie-(voorkeur) tests, moet de onderzoeker steeds zich ervan bewust zijn, dat zijn of haar persoonlijke, niet testgebonden beeld, dat hij of zij zich van de geteste heeft gevormd, niet de verwerking van de testgegevens mag doorkruisen.

Een tweede element dat de interpretatie door een beoordelaar kan beïnvloeden, is de meetbaarheid van het onderzoeksprobleem. Is het probleem niet eenduidig meetbaar, dan zal een onderzoeker ook niet eenduidig kunnen interpreteren, ofwel hij kan niet objectief zijn, zelfs al wil hij dat.

Het probleem van de objectiviteit speelde zeker in de beginperiode van de testontwikkeling een grote rol. Vandaar dat in het kader van de generaliseerbaarheid (= algemene geldigheid) van testresultaten (vooral vanuit Amerika) vele objectieve tests zijn ingevoerd.

Een objectieve test is een test die zeer welomschreven moet worden afgenomen en gescoord. De interpretatie ligt ook geheel vast. Meestal zijn het schriftelijke tests (maar dat hoeft niet per se) die door de (proef)persoon zelf worden ingevuld, bijvoorbeeld door een antwoord-alternatief aan te kruisen.

Een nadeel van objectieve tests is doorgaans het feit dat slechts een enkel (of een beperkt aantal) kenmerk(en) gemeten wordt, daar waar zo'n kenmerk beter begrepen zou kunnen worden in samenhang met andere kenmerken. (Zelfs als dit mogelijkwerijs door de onderzoeker wordt ingezien, dan nog kan hij dat niet verwerken, zijnde een subjectieve bijdrage aan de objectieve test!)

Is het niet mogelijk ten dienste van een bepaald onderzoeksprobleem een objectieve test te vinden (of te ontwikkelen), dan is men aangewezen op een test, waarbij (ongewilde) subjectiviteit van de onderzoeker een rol kan spelen. Bij classificatie- en coderingsproblemen in de verwerking en interpretatie van de testgegevens is het dan ook de gewoonte twee of meerdere 'codeurs' die gegevens te laten verwerken. Zodoende is het mogelijk de mate van overeenstemming tussen deze codeurs te berekenen. Is deze hoog, dan zal de objectiviteit van beoordeling enigszins gewaarborgd zijn. Dit noemt men de *interbeoordelaar-betrouwbaarheid* (*intercoder-agreement*, zie hfdst. 2) en het is in feite een *intersubjectiviteitsmaat*. Objectiviteit is hier dus getransformeerd tot de mate waarin beoordelaars in hun subjectiviteit overeenstemmen.

Naast het feit dat voor algemene geldigheid ten aanzien van de interpretatie van de testresultaten het gebruik van de objectieve test welhaast onontbeerlijk is, is dit soort tests wetenschappelijk gezien

belangrijk. Als een onderzoeker gebruik maakt van een objectieve test en de resultaten van zijn onderzoek in een publikatie verwerkt, dan is het voor een willekeurige andere onderzoeker mogelijk de resultaten te controleren (verifiëren), omdat hij over exact hetzelfde instrument kan beschikken. Dit geeft bij voorgeschreven gebruik dan vergelijkbare uitkomsten en interpretaties. Dit noemt men de *reproduceerbaarheid* van de test. Een ander voordeel is de *openheid* van het gebruik. Iedereen kan weten hoe je de testresultaten hebt verwerkt (geïnterpreteerd). Openheid is een interessante bijdrage tot de geloofwaardigheid van onderzoek. Als aan eisen van efficiëntie, standaardisatie, normering en objectiviteit is voldaan, dan betekent het nog niet dat een test of een test-item goed is. Bovenstaande heeft alleen te maken met regels voor afname en scoring. De volgende testvraag (item) kan aan bovenstaande eisen voldoen:

'Hoe lang was Macbeth? a. 4 liter; b. 3 kilo; c. 6 hectare; d. 2 uur.'

*In de vijfde plaats* bedenkt de raadsheer, die aan zijn aangeboren argwaan vermoedelijk het leven heeft te danken, dat zijn hoofd wel eens spoedig op een platina schaal aan hare keizerlijke gemalin zou kunnen worden aangeboden, als haar echtgenoot even argwanend is als hijzelf. Stel je voor dat de keizer geen vertrouwen heeft in de test en deze nogmaals wil laten afleggen om te zien of dezelfde 75 man overblijven. (Gelukkig bedacht de raadsheer al vrij snel dat zulks technisch onmogelijk was, zelfs als de keizer het zou willen.) Het is niet verwonderlijk dat men eist dat een test steeds hetzelfde resultaat moet opleveren. Als we van een groep kinderen het gewicht bepalen en we stellen aan de hand hiervan een rangorde op van het zwaarste tot het lichtste kind, dan zal men bij een hermeting enige tijd later (vrijwel) dezelfde rangordening moeten vinden. Dit heet *meetbetrouwbaarheid*. Onze meetlat is een betrouwbare meetlat, in dit geval de gebruikte weegschaal. Doordat we nu over twee metingen beschikken, kunnen we, door de verkregen rangordes met elkaar te vergelijken, tevens een maat vinden voor de *betrouwbaarheid* van het gebruikte instrument, bijvoorbeeld de *rangordecorrelatie*.

Mensen testen op lengte, gewicht, kledingmaat en dergelijke is niet zo moeilijk. De gebruikte tests zijn zeer eenvoudig, gestandaardiseerd, objectief en zeer betrouwbaar (bij juist gebruik). Dat ligt natuurlijk wel anders bij de in de sociale wetenschappen gebruikte tests. Alle voorgaande eisen, waaraan een test zo goed mogelijk

moet voldoen, spelen een rol bij het verkrijgen van steeds een zelfde testscore bij herhaald afnemen van de test bij dezelfde proefpersoon. Het is echter vaak erg vervelend als een testpersoon exact dezelfde items in één test tegenkomt. Ieder voor zich zal er anders op reageren: sommigen zoeken terug om te controleren of ze zo'n item wel hetzelfde beantwoord hebben als de eerste maal, anderen gaan juist iets anders invullen. Er zijn gelukkig verschillende methoden, behalve het herhaald afnemen van de test, gevonden om de betrouwbaarheid van een test te kunnen vaststellen.

*In de zesde plaats* en ten laatste bracht de raadsheer, wiens legendarische argwaan inmiddels tot ver buiten de landsgrenzen bekendheid had gekregen, enige onrustige nachten door als gevolg van een nieuwe inval: Kunnen de lijfwachten de keizer wel beschermen? Misschien heb ik wel 75 exquisite maitres de cuisine geselecteerd of misschien wel 75 gevaarlijke gekken, maar geen geschikte beschermers van zijne goddelijke majesteit. 'Heb ik met mijn test wel gemeten wat ik met de test bedoelde te meten?'

Het is vaak lastig om van een test na te gaan in hoeverre de testuitslag de juiste is, zelfs indien het een objectieve en betrouwbare test betreft. Als een test-interpretatie is: 'Deze jongeman heeft op de LOM-school de meeste kans van slagen', is dat dan ook zo? Is dat wel het geval, dan noemt men de test *valide*.

Onderzoek naar de voorspelbaarheid van de test is een wijze om te bekijken hoe valide de test is. Bijvoorbeeld: men gaat na hoeveel succes kinderen op LOM-scholen hebben, die middels de test geadviseerd zijn deze onderwijsinstelling te volgen. Ook gaat men na hoeveel succes kinderen hebben die niet door hun ouders naar de LOM-school gestuurd zijn, maar middels de test wel geadviseerd zijn dat te doen. Het criterium waarop je de test valide of niet kunt noemen, is hier de mate van succes. Dit heet de *criterium-variabele* (of het *criterium*). Men kan een maat verkrijgen om validiteit in uit te drukken door de samenhang te berekenen tussen de test-interpretatie (het advies) en het criterium (wel of geen succes, of de mate van succes).

Als men de voorspelbaarheid van het testresultaat onderzoekt, zoals bovenstaand het geval is, dan heeft men te maken met *predictieve* of *voorspellende validiteit*. Niet alle tests hebben evenwel het doel om te voorspellen. Het is dan ook zinloos de test op voorspelbaarheid te valideren.

Men kan de test laten 'concurreren' met een vergelijkbare test of

met de mening van een deskundige. Bijvoorbeeld: met behulp van een ontwikkelde diagnostische test scheidt men psychotische patiënten van gewone zenuwlijders. Nu worden de testresultaten vergeleken met de diagnose, die eerder door een psychiater is gesteld. Ook hier kan men de mate van samenhang (bijvoorbeeld de correlatie) berekenen tussen testuitslagen en de gestelde diagnoses (criterium). Zoiets heet dan: *concurrente validiteit*.

Beide vormen van validering noemt men *criterium-gebonden of pragmatische validiteit*. *Criterium-gebonden* omdat we te maken hebben met het verband tussen testuitslagen en een criterium; *pragmatisch* om aan te geven hoe bruikbaar het instrument (de test) in de praktijk is.

Soms kan het zijn, dat het kiezen van een zeker criterium slechts tot het verschuiven van de problemen leidt. Stel je voor dat de psychiater in het bovenstaande voorbeeld een zeesleper is of een belastingconsulent. Dan zal een goede validiteit tot sombere vermoedens omtrent de bruikbaarheid van de test leiden! (Tenzij de 'psychiater' evenzeer geschikt is voor dit diagnostisch werk als een echte psychiater.)

Daar waar het moeilijk is een (objectief) meetbaar criterium te vinden, is het vaak lastig een test goed te valideren.

Een derde vorm van validering mag niet ongenoemd blijven: de ophet-oog-validiteit (*face-validiteit*). Hier komt het gezonde verstand (de 'subjectiviteit') aan te pas. Construeert men bijvoorbeeld een test om het gebruiksgemak van gebitsprothesen van verkeersslachtoffertjes in een kinder-revalidatiecentrum te bekijken, dan kan daarin de vraag (test-item) voorkomen: 'Eet je wel eens toffees?' Iedere leek (en de meeste onderzoekers) zal inzien dat dit item veel te maken heeft met de onderzoeksvraag. De subjectieve *indruk* dat het item refereert aan de gestelde vraag, geeft de face-validiteit. Het item echter: 'Droom je wel eens van het sprookje van Roodkapje? Zo ja, van welke mensen of beesten uit dat sprookje dan het meest?' – heeft op het eerste gezicht niets te maken met de onderzoeksvraag. Is de onderzoeker in staat aannemelijk te maken dat het antwoord: 'Ja, over de Grote Boze Wolf', betekent dat het kunstgebitje niet lekker zit, dan heeft het item enige face-validiteit verkregen.

Face-validiteit is geen vereiste voor een meetinstrument. Indirecte tests (als de inktvlekken-test, Rorschach-test) bezitten automatisch geen face-validiteit. Bij face-validiteit gaat het erom dat je, zonder uitgebreid validatie-onderzoek, een voor een bijzondere

vraagstelling ontworpen meetinstrument lijkt te kunnen gebruiken. Als alle items intersubjectief beredeneerbaar zijn als bijdragen tot het antwoord op de vraag, dan heb je een op het oog geschikte test.

Face-validiteit heeft vaak als gevolg dat het motiverend werkt op iemand die getest wordt. De geteste heeft namelijk voortdurend in de gaten wat er bij hem getest wordt, tenzij hij denkt in het ootje genomen te worden.

Er zijn dus twee hoofdvormen van validering van een test behandeld:

a. De voorspellende validiteit, ofwel: hoe juist voorspelt de test wat ik wilde voorspellen.

b. De concurrente validiteit, ofwel: hoe zuiver werkt de test in de praktijk.

Voor beide geldt dat de correlatie berekend kan worden tussen wat de test 'zegt' en een meetbaar criterium uit de werkelijkheid.

Een 'handige' vorm van validering is de face-validiteit, ofwel: is het redelijkerwijs aannemelijk te maken dat de test zal meten wat ik wil weten? Een indirecte test kan nooit face-valide zijn. Een standaardtest (een algemeen gebruikte test) kan face-valide zijn, maar moet zeker pragmatisch valide zijn.

Binnen de testtheorie bestaan er nog andere vormen van testvalidering. Deze zijn doorgaans theoretisch interessant en praktisch van belang voor onderzoekers die zich willen bezighouden met testontwikkeling en dat zijn weer theoretici...

### *Samenvatting en generaliseerbaarheid*

Een test is een instrument om kenmerken te meten die zonder zo'n instrument niet afleesbaar zijn, of niet zo gemakkelijk, of zo goedkoop, of zo snel.

De test wordt gebruikt als bijdrage tot beantwoording van een gestelde onderzoeksvraag. Testen kan dus geen doel op zich zijn, noch kan het een instrument zijn om onderzoeksvragen op te werpen als men er om een verlegen zit. De test wordt in de sociale wetenschappen gebruikt om vragen van selectieve, diagnostische of wetenschapstheoretische aard te helpen oplossen.

Om zo goed mogelijk te kunnen meten moet de test aan een aantal eisen voldoen:

1. *Efficiëntie.* Test-items moeten zodanig geconstrueerd zijn, dat

proefpersoonsgebonden toevalsfactoren geen invloed hebben op het testresultaat.

2. *Standaardisatie*. De testafname moet 'standaard' gebeuren. Iedereen legt onder voorgeschreven, gelijke condities de test af. De test zelf moet ook standaard zijn; exact de voorgeschreven items bevatten, die op voorgeschreven wijze aan de te testen persoon worden voorgelegd.

3. *Normering*. De test is 'geschaald'. Deze schaling is geconstrueerd op de groep (populatie) waarvoor de test bedoeld is.

2 en 3 samen staan borg voor de *vergelijkbaarheid* van de test.

4. *Objectiviteit*. De test moet altijd eenduidige interpretaties geven. Doorgaans is dit te bereiken door kwantificering van de onderzoeksvraag. Hierna kan men gebruik maken van de kwantitatief te scoren objectieve test. Is kwantificering niet mogelijk of is de onderzoeksvraag te complex, dan bepaalt de mate van overeenstemming tussen verschillende beoordelaars de mate van 'objectiviteit'.

5. *Betrouwbaarheid*. De test moet steeds dezelfde uitslag opleveren als deze bij herhaling zou worden afgenomen (en de te meten kenmerken intussen niet veranderd zijn). De betrouwbaarheid kan in een getal worden uitgedrukt. Er bestaan verschillende vormen van betrouwbaarheid, die ieder bepaalde aspecten van de meting belichten.

6. *Validiteit*. De test moet zo goed mogelijk meten wat we willen weten. Er bestaan verschillende mogelijkheden tot validering van een test.

Naarmate een test beter voldoet aan de genoemde eisen, zal de *generaliseerbaarheid* van de test groter zijn. Dat wil zeggen, dat de *steekproefmeting* aan kenmerken, die verricht is in een testsituatie, een goede afspiegeling vormt van h $\ddot{o}$ e de kenmerken bij de geteste persoon in het algemeen op het gedrag van invloed zijn. Ten aanzien van voorspellende tests geldt dan dat men een juiste verwachting heeft, hoe de kenmerken in de (gedefinieerde) toekomst werkzaam zijn. Ten aanzien van diagnostische tests geldt dan dat de in de testsituatie gemeten kenmerken ook buiten de testsituatie werkzaam zijn.

Een belangrijke restrictie ten aanzien van de generaliseerbaarheid is nu gelijk genoemd: de *testsituatie*. Deze is vaak zo gekunsteld, dat het zeer wel mogelijk is, dat sommige te meten kenmerken hierdoor beïnvloed worden. En zolang de testsituatie bestaat, is het moeilijk



die invloed te bepalen! Over dit probleem stapt men vaak te gemakkelijk heen onder het motto (zo men eraan denkt): Die invloed kun je niet meten, dus bestaat hij niet.

Uitspraken als 'De test wijst uit dat...' zouden slechts met voorzichtigheid implicerend gebruikt mogen worden.

### *Nogmaals betrouwbaarheid en validiteit*

Tot nog toe is niets behandeld over de mathematische aspecten van testconstructie. Het is echter niet aannemelijk dat de meeste lezers zich in de toekomst bezig zullen houden met het construeren van tests. Men zal doorgaans, zo men ermee van doen krijgt, gebruik maken van bestaande, algemeen erkende standaardtests. Degenen echter die meer willen weten over testtheorie worden verwezen naar de hiertoe geschreven handboeken (zie de aanbevolen literatuur).

### *Soorten van betrouwbaarheidsbepaling*

De betrouwbaarheid van een test wordt – in principe – bepaald door de mate van samenhang tussen twee testafnames (dezelfde test, dezelfde proefpersonen, dezelfde testcondities) te berekenen. In de praktijk ligt de *test-hertest-methode* toch iets lastiger dan in theorie bedacht.

Een aantal storende bronnen kan bij de bepaling van de test-hertest-betrouwbaarheid een rol spelen. Men kan verstoring krijgen – waardoor de beide testafnames minder vergelijkbaar worden – door:

Niet test-gebonden veranderingen in de geteste persoon tussen beide afnames (bijvoorbeeld persoonlijkheidsontwikkeling).

Testgewenning: de proefpersoon leert van de eerste testafname.

Testslijtage: er heeft een mogelijke normverschuiving plaatsgevonden binnen de populatie waarop de test genormeerd was.

Wegvallen van proefpersonen.

Oncontroleerbare verschillen in de houding van de proefpersoon tijdens beide testafnames, voortkomend uit bijvoorbeeld: vermoeidheid, nervositeit, verliefdheid.

Van tests die niet zo 'gevoelig' zijn voor deze *invaliditeitsbronnen*, kan de betrouwbaarheid middels de test-hertest-methode worden bepaald.

Een methode om tests, die gevoelig zijn voor bronnen waarin de

tussenliggende tijdsperiode een rol speelt, op hun betrouwbaarheid te toetsen, is de *parallel-test-methode*. Men selecteert zorgvuldig steeds twee items, die hetzelfde aspect moeten meten, maar in andere woorden gesteld zijn. Voor ieder te meten aspect van een kenmerk (of kenmerken) kiest men op die manier twee items. De betrouwbaarheid wordt berekend als de mate van samenhang tussen de parallelle items.

Zo zijn er nog meer methoden voor de bepaling van de betrouwbaarheid in gebruik. De methode die gekozen wordt is afhankelijk van de vraag *welk soort* betrouwbaarheid men wenst. De soort betrouwbaarheid is afhankelijk van welke storende bronnen men beslist wenst te elimineren en welke men van ondergeschikt belang vindt.

### *Validiteit en betrouwbaarheid*

Bij betrouwbaarheidsbepaling spelen *interne* (= binnen de test) *valideringsproblemen* een rol. Bij de validiteitsbepaling van een test gaat men na in hoeverre een test correleert met de criteria in de buitenwereld (*externe validiteit*).

1. Kan een test invalide, doch betrouwbaar zijn?

2. Kan een test valide, doch onbetrouwbaar zijn?

Het probleem ten aanzien van 1. is 'slechts' dat men wat anders meet dan men wil weten, doch de meting levert iets 'constants' op. Er zijn tests die zelfs bruikbaar blijken te zijn vanwege hun grote betrouwbaarheid, maar *onzekere* validiteit: de intelligentietest. Het is niet mogelijk nauwkeurig aan te geven wat gemeten wordt.

Ten aanzien van probleem 2. geldt dat dit vrijwel onmogelijk is (in theorie zelfs onmogelijk). Immers, als je telkens wat anders meet met je test, kun je onmogelijk resultaten verkrijgen, die te toetsen zijn aan een of meer criteria. Men kan op grond hiervan zeggen dat, als een test (extern) valide is, deze automatisch betrouwbaar moet zijn. De hele kwestie van test-betrouwbaarheid en test-validiteit is feitelijk te herleiden tot valideringsproblematiek.

Betrouwbaarheid is een vorm van *interne* validiteit: de testresultaten moeten samenhang vertonen met een intern criterium. Dit interne criterium is, dat er constant hetzelfde gemeten wordt met een test. Test-validiteit is een vorm van *externe* validiteit: de (constante) testresultaten moeten samenhang vertonen met een of meer externe (buiten de experimentele situatie gestelde) criteria.

### *Wanneer moet je beslist géén test ontwikkelen?*

Sommige tests zijn overbodig. Neem als voorbeeld een selectietest voor voortgezet onderwijs, in de vijfde klas van de basisschool af te nemen. Nu blijkt de correlatie tussen testadvies en schoolhoofdadvies zeer hoog te zijn. Wat kan men nu eenvoudiger gebruiken: het advies van het schoolhoofd of het testresultaat?

### *Testconstructie. Hoe wordt een gestandaardiseerde test gemaakt?*

Summier zullen de stappen behandeld worden die men neemt bij de constructie van een geprecodeerde item-test. Dit is een test die opgebouwd is uit een (meestal groot) aantal items, met antwoordalternatieven, waarvan één (soms meerdere) alternatief 'juist' is. In deze categorie vallen de meeste intelligentietests, prestatietests en persoonlijkheidstests.

Begonnen wordt met de nauwkeurige bepaling van wat men met behulp van de test wil meten (de onderzoeksvraag). Vervolgens worden test-items bedacht, waarbij men er van uitgaat dat deze zullen bijdragen tot de beantwoording van de onderzoeksvraag. Zo ontstaat de voorlopige test. Nu worden statistische gegevens verzameld door de test op honderden proefpersonen uit te 'testen'. Men bepaalt de correlaties tussen testresultaten en itemresultaten. Als een item hoog correleert met de totale test, dan meet het item kennelijk iets van wat de hele test meet. Op grond van deze itemanalyse worden items verwijderd, herschreven (of opnieuw getekend) en nieuwe items toegevoegd. Opnieuw volgen proefafnames, totdat men tevreden is over de items waaruit de test is opgebouwd. De reden waarom een item in een test is opgenomen, is dus de goede correlatie die het item met de totale test vertoont, maar of zo'n item op zijn inhoud ook voor zo'n test geselecteerd moet worden, is vaak de vraag.

Wat te denken van het volgende test-item: 'Iets wat je in je slaap ziet is een... a. droom; b. fee; c. wens; d. angst.'<sup>1</sup>

Volgens Freud zouden 'wens' en 'angst' in aanmerking kunnen komen. Een kind met normale fantasie zal een fee zien bij gebrek aan betere keuzen. De testconstructeurs zijn echter van mening dat een 'droom zien' van elementaire vaardigheid blijk geeft, terwijl dat niet eens beschaafd Nederlands is... (Oorspronkelijk 'to see a dream' is ook geen Engels, zelfs geen Amerikaans.) Toch zal dit item wel in de test passen (zie fase 3).

De tweede fase bestaat uit het normeren (ijken) van de test. Door-

gaans omvat dit de vaststelling van definitieve score-waarden (schaalwaarden). Men bepaalt de groep waarvoor de test bedoeld is (zie blz. 116).

De derde fase is de belangrijkste fase: interne en externe (in)validiteit worden bepaald. Hiervoor dienen de externe variabelen (criteria) gevonden te worden, die afgeleid zijn uit de onderzoeksvraag. De test wordt aan deze criteria getoetst (validering). Het is zeer wel mogelijk dat de test niet valide genoeg is en terugkeer naar fase I verdient dan aanbeveling.

Als de test klaar is, treedt de laatste fase in, waarbij de handleiding wordt samengesteld, de test ten eigen behoeve wordt gebruikt, het reclamebureau is ingeschakeld en de verkoop kan beginnen.

In de handleiding zal in ieder geval vermeld moeten zijn: het doel van de test, de groep waarvoor de test bedoeld is, de instructie voor de afname, de instructie voor de testverwerking en de testinterpretatie.

Over de instructie voor de testafname moet niet te licht gedacht worden. Hierin staan alle procedures beschreven waaraan voldaan moet worden, wil de testinterpretatie zinvol zijn: Voor sommige tests is voorgeschreven hoe groot de afstand tussen testleider en de proefpersoon moet zijn, voor alle tests is de instructie aan de proefpersonen toegevoegd, die letterlijk moet worden gegeven. Het is voorgekomen dat de instructie aan de proefpersonen in een latere versie van de test gewijzigd werd, waarna de test opnieuw genormeerd moest worden.

### *Relativering ten aanzien van de betekenis van de test*

Aan testresultaten van vooral de bekende gestandaardiseerde tests wordt vaak veel betekenis gehecht. In veel personeelsadvertenties wordt vermeld dat de sollicitant zich dient te onderwerpen aan een psychologisch onderzoek. Een Amerikaans kind krijgt tegenwoordig gemiddeld drie maal per jaar een gestandaardiseerde test onder zijn neus (daartoe behoren niet schooltoetsen). Men mag aannemen dat deze tests goed gevalideerd zijn en onder optimale voorwaarden een redelijk juiste interpretatie geven.

Men gaat er misschien te gemakkelijk van uit dat standaardisatie een voldoende waarborg is voor de optimale test-voorwaarden tijdens de afname.

Dat ruis (zie blz. 114) een niet te ontkennen rol kan spelen blijkt uit een gebeurtenis op een openbare school in Boston. De leerlingen

werden op hun intelligentie getest. Over een 10-jarig meisje werd het volgende rapport uitgebracht: 'Het IQ van het kind is zo laag, dat het niet te meten is!' Aanbeveling: speciale begeleiding in de klas is noodzakelijk, beter niet op deze school. Ze werd van school genomen, maar omdat er geen andere school te vinden was, mocht ze een paar maanden later toch weer terugkeren op de oude school. Hier werd ze toevertrouwd aan een andere psycholoog, die lang en vriendelijk met haar sprak en haar daarna buiten de school opnieuw een intelligentietest afnam. Binnen een paar maanden was haar IQ van onmeetbaar laag gegroeid tot 115, waarmee ze tot de 17% slimste mensen behoort.<sup>2</sup> Kennelijk had dit meisje bij de eerste afname zo'n moeite om de test te maken (angst om te falen, angst voor de omgeving of iets dergelijks) dat ze niets presteerde.

Dit *testgedrag* wordt wel onderkend, maar het is moeilijk er rekening mee te houden bij de interpretatie, vooral omdat niet iedereen hetzelfde gedrag vertoont en omdat dit testgedrag zich (mede hierdoor) moeilijk laat meten.

Welk testgedrag het juiste is, blijkt uit de eisen die Jerrold R. Zacharias opstelde om een optimaal testresultaat op een gestandaardiseerde test te behalen:

1. Belachelijk goed kunnen lezen en schrijven.
2. Net het juiste niveau van naïviteit of sophistication hebben.
3. De taal en vooral het jargon van de tester willen begrijpen.
4. Bereid zijn te raden wat de tester nu eigenlijk wil weten.
5. Snel willen zijn.
6. De drang om te denken, iets af te maken en je zorgen te maken over een antwoord kunnen onderdrukken.

Dat het geen sinecure is voor degene die getest wordt om dit optimale gedrag te ontwikkelen, moge blijken uit hoe de items bij hem kunnen overkomen. Ze zijn vaak dubbelzinnig, betekenisloos, triviaal, misleidend, te complex wat taal betreft, uitnodigend om naar dubbele bodems te zoeken die er niet zijn of juist het omgekeerde.<sup>3</sup>

Dat een te testen persoon bereid moet zijn de drang om te denken te onderdrukken, wordt geïllustreerd aan het test-item: 'Wetenschappers houden zich bezig met drie dringen: dieren, planten en...  
*a.* mensen; *b.* sterren; *c.* mineralen; *d.* eten; *e.* godsdiensten.'<sup>4</sup>  
Iemand die aan de bovengenoemde zes eisen tegelijk wil voldoen, heeft mogelijkerwijze minder moeite met de onderstaande drie items dan iemand die even nadenkt: 'De beste verklaring voor storm en wind is: *a.* dat de aarde om zijn as draait; *b.* dat op

verschillende plaatsen verschillende temperaturen heersen; *c.* dat de aarde snel door het heelal beweegt; *d.* dat de luchtdruk van plaats tot plaats verschilt.<sup>5</sup>

Drie van de vier antwoordalternatieven hangen samen met een verklaring voor wind en storm, maar geen enkele geeft hem.

'Wat gebruiken wetenschappers om kleine dingen groter te doen lijken? *a.* barometer; *b.* lakmoespapier; *c.* weegschaal; *d.* microscoop.'<sup>5</sup>

Welke kleine dingen? Kleine luchtdrukverschillen? Kleine verschillen in zuurgraad? Kleine gewichten? Kleine objecten?

'Men hangt een natte handdoek in een warme droge kamer. Wat gebeurt er met het gewicht van de handdoek? *a.* wordt meer; *b.* wordt minder; *c.* blijft gelijk; *d.* weet niet.'<sup>6</sup>

Alternatief *d.* komt in aanmerking met de aanvulling... wat de testconstructeur bedoelt: Het gewicht van de handdoek met of zonder water?

Van de intelligentietest is bekend dat de interne validiteit meestal wel groot is (betrouwbaarheid), maar wat precies gemeten wordt, is niet immer even duidelijk. Over één zaak is iedereen het eens: intelligentie behoort los te staan van kennis. Bekijkt men dan eens dit item:

'Vogels kunnen vliegen; vleermuizen kunnen vliegen. Hieruit volgt dat: *a.* een vleermuis een vogel is; *b.* een vleermuis een zoogdier is; *c.* andere dieren dan vogels kunnen vliegen.'<sup>7</sup>

Zonder het gegeven dat vleermuizen geen vogels zijn, is er geen zinnig antwoord te geven. Tenslotte gaat het hier om logisch redeneren en niet om een biologische vraag. Om logisch te redeneren (wat volgens velen iets met intelligentie te maken heeft) mag je geen andere gegevens gebruiken dan die in de stellingen.

Zouden de beide tests, waaruit de onderstaande items afkomstig zijn, misschien aan hetzelfde criterium gevalideerd zijn?

'Wim kocht twee blocnotes à *f* 1,25 en vier pennen à *f* 0,75. Hoeveel gaf hij uit? *a.* 2,00; *b.* 3,25; *c.* 5,50; *d.* 4,25.'<sup>8</sup>

'Marie kocht een stripboek voor *f* 1,00, kauwgom voor *f* 0,25 en drop voor *f* 0,25. Hoeveel gaf ze uit? *a.* 1,25; *b.* 1,50; *c.* 1,75; *d.* 2,00; *e.* een ander bedrag.'<sup>9</sup>

Het eerste item is afkomstig uit een prestatietest, het tweede uit een intelligentietest.

Deze voorbeelden zijn allemaal uit Amerikaanse tests, maar van veel van die tests bestaan vertalingen die in Nederland worden gebruikt.

Het verregaand betekenis toekennen aan testuitslagen is misplaatst, niet alleen vanwege de vage validiteit van vooral intelligentietests en de ruis die optreedt door verschillend testgedrag, maar ook omdat een test een momentane meting geeft van karakteristieken (kenmerken) die niet blijvend behoeven te zijn. Het behalen van een 6 voor een wiskundetest betekent beslist niet, dat dezelfde persoon altijd een 6 voor wiskunde zal behalen. Zelfs IQ-scores blijken voor een persoon in de tijd zeer te variëren. Niet te vergeten de vaardigheidstests. De testresultaten hiervan hebben in de meeste gevallen een beperkte geldigheid in de tijd.

Tot slot een schrijnend voorbeeld van status-ontlening aan een IQ-score.

'In een kleine industriestad in het westen van Massachusetts zag ik op een regenachtige middag voor de deur van haar school een meisje dat me een droevig verhaal vertelde. 'Georgia Willows,' zo begon ze, 'heeft sinds de kleuterschool bij mij in de klas gezeten. Ik heb haar nooit echt aardig gevonden, maar ik moest haar wel vaak zien, omdat onze ouders met elkaar bevriend waren. Ik herinner mij dat ze in een heel groot huis woonde. Meestal speelden we daar. Georgia dacht altijd dat ze het slimste meisje van de hele wereld was. Een gave Gods weet je wel. Tegen de tijd dat we op het gymnasium zaten, beschouwde iedereen haar als het slimste meisje van de school. Toen bezocht vorig jaar, in het kader van een uitwisselingsprogramma, dat meisje uit Holland onze school. Niet alleen zag ze er verrukkelijk uit, ze was ook nog slimmer dan Georgia. Ik denk dat ze ongeveer zes talen sprak. Zelfs was ze nooit eerder in Amerika geweest en toch sprak ze de taal beter dan ik. En ze had een heleboel boeken gelezen, waar ik zelfs nog nooit van gehoord had, Amerikaanse boeken. Dus Georgia had nu werkelijk iemand om binnen de school mee te wedijveren. Ze had gedacht dat ze het op school helemaal zou maken totdat die Stephanie kwam, en Georgia raakte helemaal overstuurd.

Op een keer bij ons thuis, toen haar ouders ons bezochten, begon ze plotseling te huilen. We waren boven op mijn kamer en zo maar zonder reden begon zij te huilen. En ze bleef maar zeggen: 'Als ik niet de intelligentste van de klas ben, als Stephanies IQ hoger is dan

het mijne, dan wil ik niet meer verder leven!' Ik dacht niet dat het ernst was, maar dat was het wel. Net voor Pasen pleegde ze zelfmoord. Haar moeder vertelde me dat Georgia er niet tegenop kon om niet de knapste van de klas te zijn. Ik denk dat als het IQ zóveel voor iemand betekent, er iets helemaal verkeerd moet zijn. Maar misschien is er iets mis met het testen van mensen, waardoor ze gaan denken dat er niets belangrijkers in hun leven is dan te weten hoe knap ze wel niet zijn.<sup>2</sup>

#### *Aanbevolen literatuur*

Delen van de hieronder genoemde literatuur zullen u bekend voorkomen.

P. J. D. Drenth, *Inleiding in de testtheorie*. Deventer, Van Looghum Slaterus, 1975.

J. de Zeeuw, *Algemene psychodiagnostiek*, II. Amsterdam, Swets en Zeitlinger, 1978.

P. L. Houts (ed.), *The Myth of Measurability*. New York, Hart Publishing Comp. Inc., 1977.

#### *Noten*

1. *Iowa Test of Basic Skills*, Primary Battery, Level 7, Form 5, Vocabulary Subtest.
2. Thomas J. Cottle, 'Going up, going down', in: *The Myth of Measurability*, 1977.
3. Jerrold R. Zacharias, 'The trouble with IQ tests', idem.
4. *Scholastic Tests*; Educational Development Series, Elementary Level, Form B, 1972.
5. *Achievement Series*, Blue Green Series, 1968.
6. *Stanford Achievement Test*: Intermediate Level II, Form A, 1972.
7. *California Test of Mental Maturity*, 1963, Revision, Level 2, Test 9.
8. *Metropolitan Achievement Tests*, Elementary Level, Form J, 1970.
9. *Lorge-Thorndike Intelligence Tests*, Level 4, Form A, 1954-1966.



## 8. DE KEUZE VAN DE METHODE

Tot nu toe is alleen besproken hoe een bepaalde methode toegepast moet worden; waar je op moet letten om te zien of de resultaten van een interview de mening weergeven van de geïnterviewde en niet die van de interviewer enz. Wat nog niet aan de orde is gekomen, is wanneer je welke methode moet toepassen: wanneer ga je interviewen, observeren, testen enz. Hoewel deze keuze niet eenvoudig is en vaak zeer subjectief, zullen we proberen een paar aanwijzingen te geven.

### *De keuze van het probleem*

In de inleiding hebben wij er reeds op gewezen dat de keuze van het probleem, het omzetten van verbazing in een onderzoeksplan, een heel belangrijke stap is, die men echter zeer moeilijk aan regels kan binden. De daar genoemde factoren, zoals de wens carrière te maken en standing te verwerven, zullen nooit vermeld worden in de verslaggeving. De lezer zal deze zelf moeten opsporen door de opzet, operationalisaties, steekproeven enz. kritisch te beschouwen. Er is daarnaast nog een andere factor die evenzeer van invloed is op de uitkomsten van onderzoek, en die in ieder geval door een deel van de onderzoekers wel vermeld wordt. Deze is hoe de onderzoeker denkt dat de maatschappij en de mens in elkaar zitten. Het meest nadrukkelijk wordt dit punt naar voren gebracht door onderzoekers die door het *marxisme* geïnspireerd zijn. (Zo nadrukkelijk zelfs dat zij volgens sommigen in het geheel niet meer toekomen aan het doen van werkelijk onderzoek.) Andere mensbeelden vindt u in Van Parreren e.a. Omdat deze stroming sinds het eind van de jaren zestig, vooral onder studenten, steeds meer in de belangstelling is gekomen, willen wij er hier enige aandacht aan besteden.

De twee belangrijkste stellingen waarop de marxistische richting zich baseert zijn:

- a. Elk menselijk handelen dient gericht te zijn op de emancipatie van zichzelf en van zijn lotgenoten.
- b. Men kan nooit iets relevant te weten komen over een mens als

men zijn omgeving (inclusief de natuurlijke, maatschappelijke en zijn persoonlijke geschiedenis) er niet bij betreft.

De relatie tussen het individu en zijn omgeving is iets waar verschillende visies op bestaan. De oudste groep, die veel inspiratie uit de psychoanalyse heeft geput (men noemt ze wel freudomarxisten), ziet de maatschappij als een belemmering in de ontplooiing van het individu. Zo wordt in de 'burgerlijke' maatschappij, omdat slechts de commerciële waarde van de arbeidskracht wordt gewaardeerd en geëxploiteerd, de mens vervreemd van zijn produkt en zodoende van zichzelf als scheppend individu.

Aanhangers van de *kritische psychologie* zien de omgeving niet als alleen maar belemmerend, zij zien de ontplooiing van de mens als het resultaat van de wisselwerking tussen individualiteit en maatschappij. De invloed van de maatschappij op het individu is zeer sterk en dringt zelfs door in functies als waarnemen en denken. Wat hij uit het vele waarneemt en hoe hij dat doet, is de mens geleerd. Een mens is dan ook niet in staat los van de maatschappij te gaan staan en deze 'van bovenaf' te beschouwen. Hij draagt die maatschappij in zich.

De derde richting, de *anti-humanistische*, ziet de mens als drager van de maatschappelijke structuur. Dat mensen de illusie hebben dat ze ook nog iets anders hebben dan de kenmerken van hun maatschappelijke positie, is een door ideologie ingegeven illusie. Dat deze visies verschillende effecten hebben op de omschrijving van het begrip emancipatie spreekt vanzelf. De eerste stroming zal zich richten op het zoveel mogelijk verwijderen van alle storende invloeden van de maatschappij, opdat het individu zich volledig zal kunnen ontplooiën. Voor de laatste richting is dit een onmogelijkheid. Het individu is niets en kan dus ook niets worden zonder maatschappij.

De meeste aandacht heeft de kritische psychologie getrokken. Wat ook niet zo verwonderlijk is, want ook de onderzoeker zal niet graag zien dat hem elke individualiteit ontnomen wordt, terwijl men zich een mens onbezoedeld door de maatschappij ook moeilijk kan voorstellen.

We kunnen ons nu afvragen hoe een kritisch psycholoog een onderzoek dient op te zetten. Allereerst moeten wij ons realiseren dat die onderzoeker ook de maatschappij in zich draagt. Daar er echter individuele verschillen zijn, verschaft dit hem de mogelijkheid, in ieder geval ten dele, de maatschappij objectief te bekijken. Hij dient

dan een probleemgebied te vinden waarin de mogelijkheid aanwezig is de omgeving te bestuderen en later deze zo te veranderen dat de mensen beter tot hun recht komen, dat wil zeggen een grotere betrokkenheid bij de produkten en grotere medezeggenschap over het produktie-apparaat krijgen.

Een ogenschijnlijk goed onderzoek zoals hieronder beschreven, faalt volkomen op deze punten. In een fabriek waar veel routine-arbeid werd verricht, werden regelmatig nieuwe modellen in produktie genomen. Dit betekende dat de werknemers telkens iets andere handelingen moesten verrichten, en omdat hun loon gekoppeld was aan hun produktiviteit, ging hun inkomen achteruit. Om dit te voorkomen stelde de directie bij het invoeren van veranderingen, naar haar idee, lage normen in. Deze moesten voor de arbeiders eenvoudig overschreden kunnen worden, zodat zij alsnog aan een normaal salaris zouden kunnen komen. De normen werden echter steeds slechter geaccepteerd en de produktiviteit bleef ver achter bij de verwachting. De directie verzocht daarop een groep sociaal wetenschappelijke onderzoekers een oplossing te bedenken. Deze wetenschappers deden enige experimenten en ontwikkelden een strategie. Veel van de vervelende verschijnselen verdwenen. Hoewel de onderzoekers hier een goede kans hadden iets aan emancipatie te doen, lieten zij dat na. Zij kwamen slechts met een strategie bestaande uit een soort schijninspraak. Het zal dan ook geen verwondering wekken dat dit onderzoek voor de kritische psychologen een schoolvoorbeeld van 'burgerlijk' onderzoek is.

Ook de rol die de omgeving speelt in de ontwikkeling van het individu, heeft een invloed op het opzetten van een onderzoek. De mens is alleen zichzelf in een hem bekende omgeving. Men dient dus de mens, als men iets van hem wil weten, in zijn eigen omgeving te onderzoeken. Zo is het laboratoriumexperiment voor het verkrijgen van inzicht in de mens voor een kritische psycholoog even functioneel als een diepvrieskist op de noordpool.

Dat mensen inderdaad merkwaardige dingen doen in een laboratorium toont ook Hoogstraten aan, zij het dat hij er een andere interpretatie aan geeft (slechts de proefpersonen worden bekeken, nauwelijks de situatie). Hij legde studenten vragen voor, die onmogelijk beantwoord konden worden, zoals:

Psychologen praten in gezelschap altijd	
over hun werk	soms/meestal
Meestal heeft psychologisch onderzoek	
geen maatschappelijke relevantie	soms/altijd

Zonder enig protest of commentaar werden de vragen door vrijwel alle studenten ingevuld.

### *Conclusie*

We hebben gezien dat voor de marxistische onderzoeker niet de 'waarheid' centraal staat, maar de 'bevrijding' van de medemens. Daarnaast dat onderzoek naar de mens niet kan gebeuren zonder zijn omgeving erbij te betrekken. (Holzkamp heeft dat vergeleken met het onderzoeken van een vis zonder te weten wat water is.)

Deze vooronderstellingen zullen ongetwijfeld invloed hebben op de resultaten van onderzoek. Zonder kennis van de positie van de onderzoeker is het dan ook niet mogelijk de resultaten op hun juiste waarde te schatten. Hetzelfde geldt uiteraard voor de 'burgerlijke' onderzoeker. Zoals we in de inleiding hebben gezien, dient het onderzoek voor hen veelal om vooruit te komen binnen de bestaande maatschappelijke structuur. Door de marxisten worden zij dan ook gezien als *de* verdedigers van de status-quo. Deze burgerlijken zijn echter veel moeilijker onder één noemer te vangen (het etiket 'burgerlijk' wijzen zij uiteraard af).

Overigens zou men de indruk kunnen krijgen dat er twee soorten wetenschappers zijn: baantjesjagers en bevrijders. Dat is natuurlijk overdreven. Er zijn vele schakeringen tussen deze uitersten. Dat doet echter niets af aan het feit, dat de gebruiker van de resultaten recht heeft op informatie over het standpunt van de onderzoeker.

### *Hard of zacht*

De volgende keuze die gemaakt moet worden, is die tussen het toepassen van een harde of een zachte methode. Wij hebben die begrippen vertaald als: hypothesen toetsen of hypothesen ontwikkelen. Dit geeft al weer dat deze begrippen meer in elkaars verlengde liggen dan elkaars tegengestelde zijn. Immers, men kan moeilijk een hypothese toetsen als die er niet is, anderzijds resulteert hypothesen opwerpend onderzoek in hypothesen die vaak zeer gebonden zijn aan de onderzoeker, de onderzochte groep en de situatie. Toch woedt er al tientallen jaren een hevig gevecht tussen de voorstanders van beide benaderingen. Ieder vindt dat slechts zijn methode tot de ware kennis leidt. Uiteraard is het niet zo dat ze alleen maar tegen windmolens vechten. Het voornaamste verschil ligt in het begrip 'ware kennis'.

De 'zachten' (een vergelijking met eieren dringt zich op, maar zou

de discussie vertekenen) zijn zich wel degelijk bewust van de beperkte geldigheid van hun bevindingen, maar stellen daar tegenover dat feiten die algemeen geldig zijn ofwel triviaal ofwel irrelevant zijn. Zij zullen zich dan ook zelden bezondigen aan uitspraken over 'de mens' en zich beperken tot de studie van één specifiek object. Ook hun conclusies en aanwijzingen ter verandering of verbetering hebben slechts geldigheid voor dat ene geval. Het is echter niet zo, dat hun bevindingen in die specifieke situatie geheel nutteloos zijn in andere situaties. Elk mens, elke groep, elk instituut verschilt van de ander op een aantal punten, maar heeft ook veel gemeenschappelijks. Men hoeft slechts na te gaan waarin ze verschillen om ook in een andere situatie zinvol werk te kunnen verrichten.

De 'harden' redeneren min of meer andersom. Zij willen in de eerste plaats weten wat een categorie objecten (mensen, ziekenhuizen enz.) gemeenschappelijk heeft. Welk object je dan ook later tegenkomt uit die categorie, een aantal basisgegevens heb je. Dat er vaak aanvullende informatie nodig is om conclusies te trekken of beleidlijnen op te stellen, doet niets af aan de geldigheid van de verworven kennis.

Beide methoden kunnen dus bruikbare informatie opleveren. Welke methode men kiest, is een kwestie van smaak. Nogmaals willen we er op wijzen dat de methoden bruikbare informatie op *kunnen* leveren, maar dat dat lang niet altijd het geval is. In het bijzonder geldt dit voor de harde methoden. De botte afwijzing van zacht onderzoek maakt het veelal onmogelijk relevante hypothesen op te stellen. Men kan nog zulke mooie en verantwoorde tests, observatieschema's en vragenlijsten maken, als de relevante variabelen niet opgenomen zijn, is het onderzoek doelloos en zinloos.

### *Vrij of gestructureerd*

Een heel andere vraag is of men een vrije of gestructureerde methode zal toepassen. *Vrij* betekent dat men de betreffende persoon informatie laat geven over een bepaald onderwerp (door praten of doen) zonder dat men hem op enigerlei wijze beïnvloedt. *Gestructureerd* daarentegen houdt in dat men ook na het 'startschot' zich voortdurend met de gang van zaken blijft bemoeien. Men stelt van te voren vastgelegde vragen of geeft dito opdrachten. Ook zal men herhaaldelijk bijsturen of onderbreken. Zeer gestructureerde methoden zijn bijvoorbeeld intelligentietests en multiple choice-tenta-

mens. Een voorbeeld van een vrije methode is het diepte-interview. Dit is een vaak 100 of meer uren durend gesprek waarbij alleen het onderwerp vaststaat, en dat is veelal de geïnterviewde (zie en hoor bijvoorbeeld: 'Tony van Verre ontmoette Ko van Dijk'). Tussen deze twee uitersten is een heel scala van overgangen. Binnen een interview schept men bijvoorbeeld vaak de mogelijkheid tot doorvragen (na de vraag staat dan D). Meestal gebeurt dit als men niet weet wat alle mogelijke antwoorden zijn (zodat men ze niet kan precoderen) of als er de mogelijkheid is dat de vraag niet goed begrepen wordt. Zo beantwoordde iemand op de radio de vraag 'Heeft u een donkere kamer?' met 'Ja, behoorlijk.' Ook de vrije methoden zijn over het algemeen niet volledig vrij. Meestal zijn er beperkingen in de tijd en heeft de interviewer toch een wat beperkter doel in zijn hoofd dan alles te weten te komen.

Een keuze tussen deze methoden is echter voorbehouden aan hen die een 'hard' onderzoek hebben gepland. Iemand die een 'zacht' onderzoek doet, kan onmogelijk gestructureerd te werk gaan. Hij weet immers niet wat hij allemaal zal tegenkomen.

Een harde methode kan wel gecombineerd worden met een vrije manier van gegevens verzamelen. Laten we als voorbeeld nemen dat we de leeftijd van iemand te weten willen komen. Over het algemeen zal dat weinig problemen opleveren. Er zijn echter mensen, vrouwen niet uitgezonderd, die daar een geheim van wensen te maken. De oudere diva zal aanzienlijk jonger zijn, als we haar er gewoon naar vragen, de prille puber op een brommer zal de andere kant op liegen. Bij het toepassen van een vrije methode zal het veel moeilijker zijn. Veel gebeurtenissen zijn leeftijd- en tijd-gebonden. Als iemand bijvoorbeeld vertelt dat hij als dienstplichtig militair geholpen heeft bij de watersnoodramp, dan was hij dus 19 of 20 in 1953.

Een mooi voorbeeld vinden we bij Bleek (*Anders bekeken*), die in Ghana informatie verzamelde over de kennis van methoden van geboortenbeperking. Links staan de gegevens verkregen met behulp van een standaard vragenlijst, rechts verzameld met een vrije methode (participerende observatie). Beide methoden werden uiteraard niet door dezelfde onderzoeker toegepast.

*Vragenlijst*

24 jaar oud

éénmaal gescheiden

vier kinderen gehad

viermaal zwanger geweest

haar dienstmeisje maakt het  
eten klaar

woont met echtgenoot in Accra

heeft nooit enige vorm van  
geboortenbeperking toegepast

nooit abortus ondergaan

*Participerende observatie*

31 jaar oud

tweemaal gescheiden

zes kinderen gebaard  
(2 overleden)minstens negenmaal  
zwanger geweest

heeft geen dienstmeisje

woont alleen in Ayere

heeft ruime ervaring met  
allerlei methodenminstens driemaal abortus  
ondergaan

Door ook op andere manieren deze resultaten te controleren bleek de vrije methode superieur. Er zijn echter ook voorbeelden van het omgekeerde te vinden. Zo kwamen twee onderzoekers, die onafhankelijk van elkaar een verslag schreven van hun vrije observatie van hetzelfde dorpje, Tepotzlan in Mexico, tot extreem verschillende omschrijvingen. De een vond het triest en naargeestig, de ander vriendelijk en pittoresk. De resultaten van gestructureerde onderzoeken over hetzelfde object zullen nooit zo ver uiteen liggen.

We zien hieruit dat een belangrijk verschil tussen beide werkwijzen het accent is dat ze leggen op validiteit dan wel betrouwbaarheid. De vrije methode legt sterk de nadruk op validiteit, maar omdat de kans groot is dat een andere onderzoeker met andere resultaten komt (volgens sommigen is dat per definitie zo bij de vrije aanpak), kan men moeilijk beslissen welke van deze de ware is.

Aan de andere kant geeft de hoge betrouwbaarheid, waar de gestructureerde methode naar streeft, geen enkele zekerheid betreffende het waarheidsgehalte van de resultaten (zie hoofdstuk 7).

Deze keuze lijkt dus een keuze tussen twee kwaden. Toch valt dat wel mee. In de eerste plaats zijn mensen natuurlijk niet over alles systematisch leugenachtig. Op heel veel vragen zullen ze een eerlijk antwoord geven. In de tweede plaats is er niets op tegen om de beide methoden te combineren, er zijn zelfs alleen maar voordelen, behalve dat het onderzoek meer tijd gaat kosten natuurlijk.

De laatste tijd zijn er nog twee methoden ontwikkeld die moeilijk in deze tweedeling zijn onder te brengen. Volgens de voorstanders

verenigen zij de voordelen van beide methoden zonder hun nadelen te hebben.

De eerste is de zogenaamde niets versturende meting (*unobtrusive measurement*, Webb 1966). Deze opzet houdt in dat er geen enkel contact is tussen de onderzoeker en de onderzochte. De reeds behandelde inhoudsanalyse op bestaand materiaal is er een voorbeeld van, maar ook observatie leent zich hier goed voor. Zo heeft men in Amerika onderzocht of de mensen in steden, waar in grote mate gestemd was op *law and order* partijen, zich daar ook gedroegen overeenkomstig hetgeen deze partijen voorstonden. Daartoe observeerde de onderzoeker, op een verborgen plaats, het gedrag van voetgangers bij door stoplichten beveiligde oversteekplaatsen. Er bleken inderdaad veel minder overtredingen gemaakt te worden in bovengeschetste steden dan in steden waar anders gestemd was. De andere methode is het levensechte experiment. Hierbij is er wel contact tussen de onderzoeker (of zijn medewerker) en de onderzochte, maar zonder dat de onderzochte weet dat hij deel uitmaakt van een onderzoek. Een voorbeeld hiervan vinden we bij het onderzoek van Brunt naar rassendiscriminatie op de arbeidsmarkt. Hij stuurde drie handlangers (een blanke Nederlander, een Creoolse Nederlander en een Spanjaard) naar een aantal instellingen om te solliciteren op een daar aanwezige vacature. Behalve het bovenvermelde verschil waren zij zoveel mogelijk gelijk: even oud, gelijk gekleed en zeiden dezelfde opleiding te hebben genoten. Toch bleek de blanke veel vaker te worden aangenomen dan de Creool. We maken onderscheid tussen deze methoden, in tegenstelling tot andere, omdat in het levensechte experiment moedwillige veranderingen in de omgeving van de onderzochte worden aangebracht, die niet natuurlijk behoeven te zijn. Zo is een van de punten van kritiek op dit onderzoek dat de neger van de drie niet het minst werd aangenomen omdat hij zwart was, maar omdat het een rare neger was. Hij was te netjes, te goed opgeleid enz. (In hoeverre hier racisme wordt gebruikt om racisme weg te redeneren laten wij aan de beoordeling van de lezer over.)

Nadelen van beide methoden zijn dat ze slechts voor een zeer beperkt aantal problemen toepasbaar zijn en dat de manier van meten vaak ver van het te meten begrip afstaat. Webb noemt een zeer groot aantal vindingrijke opzetten, maar ondanks dat blijken de resultaten zeer vaak voor velerlei uitleg vatbaar te zijn.



## Conclusie

De keuze van een probleem blijft een uiterst subjectieve zaak, die echter ernstige gevolgen kan hebben ten opzichte van de bruikbaarheid van de onderzoeksresultaten. Om uit deze impasse te komen lijkt het ons zeker niet verstandig een soort wetenschappelijke dictatuur in te stellen en iedereen te verplichten bijvoorbeeld de kritische psychologie aan te hangen (iedereen overtuigen zal zeker niet lukken). Onzes inziens kan hier alleen wat aan gedaan worden door meer openheid te eisen.

De volgende keuze tussen hard en zacht en vrij en gestructureerd laat zich in het volgende (verre van onfeilbare) schema onderbrengen:

1. Men kent de relevante variabelen niet of niet voldoende. Of men wenst slechts één object te onderzoeken, zonder pretenties dat dit een representant zal zijn van een grotere verzameling.

Aanpak: zacht en vrij.

2. Men beschikt wel over informatie over de relevante variabelen met betrekking tot het onderwerp, maar de kans is groot dat de meting een vertekend beeld zal opleveren.

Aanpak: hard en vrij (onder bepaalde voorwaarden *unobtrusive measurement* of levensecht experiment).

3. Men beschikt wel over voldoende informatie en men kan de variabelen eenvoudig meten of men heeft maatregelen genomen om vertekening te voorkomen (projectieve tests, indirecte vragen).

Aanpak: hard en gestructureerd.

## De keuze tussen de verzamelingswijzen

We hebben er reeds eerder op gewezen dat het vaak raadzaam is meer dan één werkwijze te gebruiken (interview + inhoudsanalyse bijvoorbeeld als het over de aankoop van huishoudelijke apparaten gaat, of test + observatie bij kinderen met bepaalde gedragsstoornissen). Het zal dan ook geen verwondering wekken dat ook hier een onfeilbaar schema zal ontbreken, waar men aan de ene kant het probleem in kan stoppen en waar aan de andere kant een pasklare methode uitrolt. We kunnen slechts een aantal factoren noemen die enige richting geven aan de keuze.

### 1. Zeggen of doen

Uit allerlei uitdrukkingen zoals: zeggen en doen is twee, de daad bij

het woord voegen en een grote mond hebben maar een klein hartje, geen woorden maar daden, blijkt dat het ook in de dagelijkse omgang bekend is dat zeggen en doen niet hetzelfde zijn. (We hebben dat ook al gezien in het onderzoek van Lapière over het Chinese echtpaar; zie hoofdstuk 3). Het doen kunnen we meten met observatie en gedragstests, het zeggen met inhoudsanalyse, vraaggesprek, vragenlijsten, schalen en de andere soorten tests. Het komt echter vaak voor dat men via het meten van 'zeggen' iets over gedrag te weten wil komen. Ook het omgekeerde kan voorkomen. Wanneer moet je nu die indirecte weg bewandelen?

In de eerste plaats zo weinig mogelijk. Als het even kan, moet je gedrag meten als je iets over gedrag wilt zeggen en moet je mensen interviewen als je wilt weten wat ze te zeggen hebben. Eén geval waarin het wel toelaatbaar is, is als je de zekerheid hebt dat in dat speciale geval zeggen en doen wel hetzelfde zijn. Stel dat je bijvoorbeeld wilt weten of iemand die een nieuw boek koopt dat ook binnen pak weg twee weken leest. Je zou dan, vergezeld van een cohort detectives, bij een boekwinkel kunnen gaan staan en aan ieder die met een boek de winkel verlaat, zo'n speurneus kunnen toevoegen. Deze blijft de persoon schaduwen tot de twee weken voorbij zijn of tot hij geconstateerd heeft dat er gedurende voldoende tijd met de neus in het boek is gezeten. Een uiterst ingewikkelde en kostbare onderneming. Veel eenvoudiger is om ze er na veertien dagen domweg naar te vragen (en hem eventueel nog te controleren door vragen over de inhoud te stellen).

Een tweede geval waarin de omweg de enige weg is, doet zich voor als bijvoorbeeld het gedrag niet direct observeerbaar is. Dat is onder andere het geval bij verkiezingen. Het is zelfs strafbaar hier direct te observeren. Als men er toch iets over wil weten, zal men dat aan de stemmers moeten vragen, of zoals vaker gebeurt, het op verzoek nogmaals laten doen, waarbij het stemformulier niet in de stembus maar in de zak van de onderzoeker verdwijnt. We vervangen observatie dan door inhoudsanalyse. Het kan ook voorkomen dat mensen over iets niets willen of kunnen zeggen. Soms is het mogelijk uit hun gedrag af te leiden wat ze gezegd zouden hebben hadden ze wel willen of kunnen antwoorden. Zo werd er verteld dat kardinaal Willebrands de nieuwe paus zou worden omdat hij met Pasen naast paus Paulus VI op het balkon was verschenen. Tot tweemaal toe werd hij blijkbaar gepasseerd.

Een probleem bij dit 'indirecte' meten blijft dat men er lang niet altijd zeker van kan zijn, dat men via die omweg toch bij het juiste

doel terechtkomt. Daarom dient men, vooral in dit geval, uiterst zorgvuldig de validiteit te bepalen.

## 2. *De plaatsing in de tijd*

Over de plaatsing in de tijd kunnen we gelukkig iets duidelijker zijn. Als we willen weten hoe iets of iemand was in het verleden, vallen automatisch observatie en tests uit. Men kan zoeken naar overblijfselen uit het verleden en inhoudsanalyse toepassen, of het de betreffende persoon of personen mondeling of schriftelijk vragen. De inhoudsanalyse heeft, indien men voldoende materiaal te pakken kan krijgen, hierbij de voorkeur, omdat vraaggesprekken en vragenlijsten beïnvloed kunnen worden door vertekening en vergeten.

Het enige ons bekende geval waarin toch observatie wordt gebruikt, is de reconstructie (vooral populair bij de politie, althans volgens schrijvers van detectives). Men laat iedereen nog eens precies hetzelfde doen als op dat 'fatale' ogenblik. Dat ook hier weer geen zekerheid bestaat dat alles inderdaad hetzelfde zal zijn, spreekt voor zichzelf. Men kan dit enigszins opvangen door vooraf alle betrokkenen te vragen hoe iedereen zich volgens hen op dat ogenblik gedroeg.

## 3. *De te meten variabele*

We hopen niet dat u verwacht dat er nu een lijst zal volgen van alle variabelen die ooit binnen sociaal-wetenschappelijk onderzoek zijn gebruikt met daarachter de best passende methode om die te meten. We zullen slechts één kenmerk behandelen, namelijk de complexiteit van de te meten variabele.

Over het algemeen kunnen we met observatie en de vragenlijst alleen maar tamelijk eenvoudige variabelen meten. Schalen en vooral tests daarentegen zijn ontworpen om zeer complexe en moeilijk te vangen variabelen meetbaar te maken. (Sir Cyril Burt was het hier overigens niet mee eens en corrigeerde uitslagen van zijn intelligentietest als deze niet overeenstemden met zijn eigen indruk van die persoon; Vroon, 1979.)

## 4. *Contact met de proefpersoon*

Bij het afnemen van een test en het houden van een vraaggesprek is contact met de proefpersoon onvermijdelijk. Dat heeft aan de ene kant voordelen: men beheerst de situatie, kan voortdurend bijsturen, de situatie verhelderen enz. Nadelen zijn dat de situatie waarin

het materiaal verzameld wordt, vaak een wat gedwongen karakter heeft en dat het contact tussen onderzoeker en proefpersoon de reacties van deze laatste kan beïnvloeden.

Voor inhoudsanalyse op bestaand materiaal geldt min of meer het omgekeerde. De onderzoeker heeft geen enkele vertekende invloed op het materiaal, maar hij kan ook nooit om verduidelijking of aanvulling vragen.

De vragenlijst heeft bijna alle nadelen van beide eerder genoemde methoden, zij het in mindere mate. Weliswaar is de onderzoeker niet persoonlijk aanwezig, maar hij wordt gerepresenteerd door zijn vragenlijst. Zowel vorm, kleur en verzorging van de vragenlijst als de toon van de vragen kunnen de antwoorden beïnvloeden.

Over observatie is minder eenvoudig een oordeel te geven. Over het algemeen heeft men de situatie wat minder in de hand dan bij de test (soms in het geheel niet). Het contact tussen onderzoeker en geobserveerde kan heel intensief zijn, maar ook totaal afwezig.

De keuze van de methode is in dit geval sterk afhankelijk van de mate waarin men denkt dat de verschillende nadelen effect zullen hebben op de resultaten. Daarnaast zal men proberen de altijd toch aanwezige nadelen te neutraliseren (bijvoorbeeld door een gestandaardiseerde testafname) of er gebruik van te maken (zie hoofdstuk 3).

##### 5. *De beschikbare tijd*

De hoeveelheid tijd die men kwijt is aan het verzamelen van de gegevens, is niet alleen van de gekozen methode afhankelijk maar ook van het aantal personen of objecten dat men wenst te onderzoeken en het aantal variabelen dat men gaat meten.

Dé methode om aan veel personen veel variabelen binnen korte tijd te meten is de vragenlijst. Voor elke proefpersoon meer hoeft men alleen maar een extra adres te schrijven.

Inhoudsanalyse zal over het algemeen meer tijd vragen. Het zoeken en selecteren van het materiaal kan soms zeer tijdrovend zijn.

Het vraaggesprek en observatie kosten weer meer tijd. Per persoon is men een aanzienlijke hoeveelheid tijd kwijt, variërend van vijftien minuten tot enkele uren.

De test is vermoedelijk het meest tijdrovend. Het afnemen van een gemiddelde intelligentietest duurt gewoonlijk twee tot drie uur.

De test heeft daarentegen wel het voordeel dat men tamelijk snel de totaal-score kan berekenen. Bij alle andere methoden kan op het verzamelen nog een zeer langdurige periode van categoriseren vol-

gen voordat men analyseerbare resultaten heeft. Men dient zich hierbij wel te realiseren dat de factor tijd alleen een rol mag spelen bij de keuze tussen twee gelijkwaardige methoden. Als men niet voldoende tijd heeft om een geschikte methode toe te passen, kan men beter van het onderzoek afzien dan zijn tijd te verdoen met het verzamelen van nonsens.

#### *Aanbevolen literatuur*

Delen van de hieronder genoemde literatuur zullen u bekend voorkomen.

- L. Brunt (red.), *Anders bekeken*. Meppel, Boom, 1977.  
F. Bovenkerk (red.), *Omdat zij anders zijn*. Meppel, Boom, 1978.  
E. J. Webb, D. T. Campbell, R. D. Schwartz & L. Sechrest, *Unobtrusive measures. Nonreactive research in the social sciences*. Chicago, Rand McNally, 1966.  
Joh. Hoogstraten, *De machteloze onderzoeker*. Meppel, Boom, 1979.  
C. F. van Parreren & J. G. van der Bend (red.), *Psychologie en mensbeeld*. Baarn, Ambo, Basisboeken, 1979.  
P. A. Vroon, *Intelligentie*. Baarn, Ambo, Basisboeken, 1979.  
Tony van Verre ontmoette Ko van Dijk. Bussum, Unieboek, 1978.

## 9. KIJKEN NAAR FEITEN: EEN ALTERNATIEVE ONDERZOEKSMETHODE

Een vrij ongebruikelijke wijze van theorievorming ten behoeve van verkennend en praktisch gericht onderzoek in de sociale wetenschappen is het opwerpen van hypothesen en theorieën *nadat* gegevens verzameld zijn. Om het bijzondere karakter van het afleiden van wetenschappelijke uitspraken uit directe feiten te begrijpen, is het misschien handig nog eens te kijken hoe de overheersende (neopositivistische) wetenschapsopvatting, met zijn empirisch of kwantitatief onderzoek, regeert.

De wetenschap probeert verschijnselen uit de werkelijkheid te verklaren, zodat ze voorspelbaar zullen zijn en zo mogelijk beheerst kunnen worden. Een verschijnsel is verklaard als aangegeven kan worden onder welke, wel-omschreven, omstandigheden het verschijnsel zal optreden en onder welke omstandigheden niet. Treedt een verschijnsel 'toevallig' op, d.w.z. niet in overeenstemming met de voorspelling, dan wil dit zeggen dat de omstandigheden, waaronder het verschijnsel zich manifesteert, niet voldoende omschreven zijn, en is het verschijnsel niet verklaard. Kent men de omstandigheden (het complex van samenhangende verschijnselen = de voorwaarden), dan zal men voorspellen dat het verschijnsel optreedt als:

1. de voorwaarden zich voordoen (theoretisch interessant) of
2. men de voorwaarden creëert (praktisch relevant).

Creëert men de voorwaarden, dan beheerst men het verschijnsel. Een voorbeeld: Agressie kan men verklaren uit frustratie (deze verschijnselen behoeven op zich eveneens een verklaring, maar daar gaat het hier niet om). Frustreert men voetballers door ze onophoudelijk onderuit te schoppen of belachelijk te maken, dan kan men een stevige partij voetbal verwachten, waarbij sterk meelevende supporters zich ook niet onbetuigd zullen laten.

De kern van wetenschappelijk onderzoek ligt vooral in het zoeken naar beheersbare (manipuleerbare) voorwaarden waaronder een gewenst verschijnsel zich voordoet (of een ongewenst verschijnsel juist niet). Nu is deze opvatting over wetenschap al bijna evenveel jaren oud, als er jaren verstreken zijn sinds het ogenblik waarop

mensen zich over verschijnselen gingen verbazen. Dit leidde ertoe dat dagelijkse verschijnselen reeds lang 'verklaard' zijn: Een spin in de morgen brengt kommer en zorgen; Bier op wijn geeft venijn; Wie wind zaait, zal storm oogsten. Sommige verschijnselen wachten nog op nadere verklarende voorwaarden: Eén zwaluw maakt nog geen zomer. Volgens sommigen brachten en brengen volkswijsheden en andere minder zorgvuldig onderzochte uitspraken de 'waarheid' niet altijd aan het licht. Dientengevolge splitst de wetenschap zich af van het dagelijkse waarnemen, beschouwen en concluderen en gaat over tot een *systematische* wijze van beschrijven van verschijnselen in samenhang met andere verschijnselen. (Zie hoofdstuk 2 voor een verdere uitwerking.)

De empirische wetenschapsbeoefening heeft een heel scala van formele handelingen geïnstitutionaliseerd. Al deze handelingen zijn bedoeld om vergissingen uit te sluiten en maken dat een wetenschappelijke uitspraak objectief is en universeel geldig. Objectief wil zeggen: gebaseerd op feiten, niet op meningen, zodat het onderzoek gecontroleerd kan worden. Als een uitspraak universeel geldig is, betekent het dat het te verklaren verschijnsel zich onder de beschreven voorwaarden wáár en wanneer dan ook zal voordoen.

(De uitspraak: 'God bestaat' is voor gelovigen waarschijnlijk universeel geldig, doch is niet gebaseerd op objectief waarneembare feiten en daarom niet wetenschappelijk.)

Sociaal-wetenschappelijk onderzoek is voornamelijk 'hypothese-toetsend-onderzoek', waarbij statistische uitspraken gedaan worden over theorieën. Een statistische uitspraak betekent dat er een zekere kans is, door onderzoeker en aard van het onderzoek bepaald, dat de uitspraak niet waar is.

De navolgende procedure zou in zijn algemeenheid het proces van dit soort onderzoek kunnen beschrijven.

1. Een onderzoeker *constateert* een niet-verklaard verschijnsel.
2. Hij trekt zich terug om te bedenken hoe het verschijnsel te verklaren is. Uit eerdere onderzoeken op deelgebieden die samenhangen met het verschijnsel, eigen creativiteit, inventiviteit, interesse, geld, tijd, geduld produceert hij een aantal vooronderstellingen, waarbij het verschijnsel in mootjes wordt gehakt. Verklaring van de elementen (mootjes) waaruit volgens de onderzoeker het verschijnsel is opgebouwd, geeft dan een verklaring voor het totale verschijnsel. Er wordt een *systematische ordening* aangebracht in de (onderzoeks)elementen op grond van logica, intuïtie en

vooronderstellingen. Het gevolg is een logische constructie die regels bevat om het verschijnsel te kunnen onderzoeken. Dit speculatieve ei noemt men:

3. De *theorie*. Een theorie is een geheel van vooronderstellingen die een verklaring geven voor het verschijnsel en waaruit verschillende uitspraken kunnen worden afgeleid (deductie) die in principe controleerbaar (toetsbaar) zijn. Blijkt een uit de theorie afgeleide uitspraak (hypothese) onwaar, dan verdient de theorie herziening.

4. Omdat de werkelijkheid zich niet zo eenvoudig leent voor controle (toetsing), moet deze meestal nagebootst worden: het Waterloopkundig Laboratorium in de N.O.-polder is een aardig voorbeeld van nabootsing van de Deltawerken.

Een representatieve steekproef uit de Nederlandse bevolking is een nabootsing van het Nederlandse volk.

De uitkomst van de formule  $S_t = V_0 t + \frac{1}{2} a t^2$  is een goede nabootsing van de werkelijke afstand die een bewegend object na  $t$  seconden zou hebben afgelegd. Zo'n hanteerbare nabootsing van de te onderzoeken werkelijkheid noemt men een *model* van de werkelijkheid. Uit de theorie afgeleide hypothesen toetst men aan het model.

5. Statistische toetsing vereist dat hypothesen kwantitatief (= meetbaar) zijn. Dit wordt een *operationalisatie* genoemd. Men kan zich misschien voorstellen dat een hypothese als: 'taakgerichtheid leidt tot grotere prestaties' lastig te meten is. 'Grotere prestaties' is te meten door het vergelijken van bijvoorbeeld de cijfers waarin de prestatie wordt uitgedrukt. Het begrip 'taakgerichtheid' laat zich moeilijker in getallen of categorieën uitdrukken. Een test (zie hoofdstuk 7) zal hier misschien uitkomst moeten bieden om het kwalitatieve begrip 'taakgerichtheid' in een getal uit te drukken. De geoperationaliseerde hypothese luidt dan: 'Hogere cijfers op de taakgerichtheidstest leiden tot hogere prestatiecijfers.' (Dat zo'n cijfer niet het gehele begrip kan vertegenwoordigen is een noodzakelijk offer aan de kwantificatie.)

6. De toetsing kan beginnen. De in aanmerking komende statistische toetsen (of andere technieken) worden uit de kast gehaald en afhankelijk daarvan worden nu *gegevens verzameld* (uit de werkelijkheid!) die nodig zijn om de theorie aan het model te toetsen.

7. Het al dan niet verwerpen van geoperationaliseerde hypothesen moet teruggekoppeld worden naar de uit de theorie afgeleide uitspraken (*interpretatie*).

8. De consequentie is, dat de theorie moet worden bijgesteld, zodat



we weer in fase 2 zijn aangeland. Het *proces herhaalt* zich totdat òf de onderzoeker inziet dat zijn theorie niet klopt (komt zelden voor), òf zoveel offers aan de kwantificatie heeft gebracht, dat het oorspronkelijke verschijnsel dat zijn verbazing opwekte, gereduceerd is tot een schijnfenomeen. (Het verschijnsel 'angst' is te verklaren als een omschreven biochemisch proces. Een verklaring waar slechts weinigen op zullen wachten.)

9. Op een gegeven moment kan de onderzoeker zijn theorie niet *falsifiëren*. Dit gebeurt als voortgezette toetsing geen aanleiding meer geeft tot het bijstellen van de theorie. Gaarne geeft men nu de theorie het voordeel van de twijfel. Zij geldt als 'waar', totdat nieuw materiaal wordt aangedragen om haar alsnog om zeep te helpen.

Er is een aantal momenten in deze benadering die nadere aandacht verdienen.

Ten eerste: Vele kwalitatieve eigenschappen die toch deel uitmaken van het onderzoeksverschijnsel, gaan bij operationalisering verloren. Vaak leidt voortdurende kwantificatie tot de verklaring van 'schijnfenomenen'.

Ten tweede: Ten aanzien van het model worden aannamen gedaan die dikwijls moeilijk te rechtvaardigen zijn. De vraag: 'Gedraagt het model zich net zo als de werkelijkheid' wordt zelden gesteld, vooral niet als er met algemeen geaccepteerde modellen wordt gewerkt. Het gevolg van het gebruik van een verkeerd model zal zijn, dat de interpretatie van de getoetste uitspraken slechts geldigheid heeft ten aanzien van het model zelf en niet zal gelden voor de werkelijkheid. Dit is het gevreesde generaliseerbaarheidsprobleem: Als uit mijn onderzoek volgt dat A de oorzaak is van B, zal buiten mijn onderzoek A dan ook de veroorzaker van B zijn?

Ten derde: Een theorie wordt a priori opgesteld (eerst theorie, dan kijken of de werkelijkheid zich zo gedraagt). Een onderzoeker zal al gauw geneigd zijn een voorkeur te benadrukken voor één bepaalde invalshoek om het te onderzoeken verschijnsel te beschouwen (zie de inleiding). De theorie zou zodoende wel eens te beperkt kunnen zijn om het verschijnsel naar bevrediging te verklaren.

Ten vierde: Bij het toetsen van hypothesen neemt men een zekere kans op een verkeerde conclusie. Daar waar meerdere hypothesen getoetst worden, zal de kans, dat ten minste een der conclusies onjuist is, toenemen.

Ten vijfde: Bij het verzamelen van gegevens stuit men vaak op

allerlei eigenaardigheden, waarvan men denkt: 'Hé, wat is dat nou?' Omdat de gegevens zeer selectief, ten dienste van de statistische toetsing, verzameld worden, kunnen dergelijke eigenaardigheden formeel niet verwerkt worden (pas als de theorie bijgesteld dient te worden, kan men er verdere aandacht aan besteden).

Enkele voorbeelden: Het valt op dat een aantal proefpersonen bij het invullen van een vragenlijst zijn zenuwen duidelijk niet de baas is. Een aantal proefpersonen schrijft hout-snijdende kritiek bij test-items. De rubriek 'geen mening' wordt nogal eens ingevuld door personen die verklaren de vragen niet begrepen te hebben.

Een aantal verschijnselen dat zich niet laat ontleden door empirisch onderzoek, zijn de individuele processen (bijvoorbeeld: gevoelservaringen en -uitingen). Het onderzoek 'bepert' zich tot universele verschijnselen, waardoor individuele ervaringen op andere wijzen dienen te worden beschreven (romans). Ook een aantal sociale verschijnselen laat zich in, als wezenlijk gevoelde, aspecten minder goed onderzoeken met behulp van een kwantitatieve methode: emancipatie, jeugdwerkloosheid, fascisme, moederschap, fobieën, stervensbegeleiding, etc.

Dat andere niet kwantitatieve methoden van onderzoek minder aandacht krijgen, is te verklaren uit een gebrek aan controleerbaarheid en een beperktheid ten aanzien van de uitspraken die deze methoden opleveren. Doordat er meestal geen sprake is van doorgevoerde kwantificering, zullen de uitspraken ook niet objectief genoeg geacht worden. ('Niet wetenschappelijk' is in onderzoeksinstellingen een genadeloze beschuldiging.) Eén van die methoden is echter interessant genoeg om bij stil te staan, omdat deze onderzoeksmethode gefundeerd is op datgene waar dit boek over handelt: het verzamelen van gegevens.

### *Funderend onderzoek*

Het doel van deze onderzoeksmethode is om uit systematisch verzamelde gegevens een verklaring te ontwikkelen voor een concreet te onderzoeken verschijnsel en deze verklaring te veralgemeniseren tot een universeel geldige theorie over het verschijnsel.

Men doet bijvoorbeeld een onderzoek naar schuldgevoelens bij mensen die aan een depressie lijden en komt uiteindelijk met een theorie over schuldgevoelens.

Zo'n gefundeerde theorie is 'niet wetenschappelijk', want zij is niet objectief toetsbaar...

Zoals gezegd is het doel om rechtstreeks uit gegevens een theorie te ontwikkelen. Dit gebeurt met een techniek die genoemd wordt: *de constante vergelijkende analyse*. Om te beginnen bedenkt men een aantal *categorieën*, waarvan men aanneemt dat ze bijdragen tot de beschrijving van het onderzochte verschijnsel.

Nu worden middels observatie, vraagg gesprekken of inhoudsanalyse kwalitatieve gegevens verzameld. Men kijkt in welke categorie of categorieën deze gegevens passen. Gelijktijdig worden nieuwe gegevens verzameld en vergeleken met de eerder verkregen gegevens. Zodoende worden overeenkomsten en verschillen waargenomen, die eigenschappen bepalen binnen de categorieën, waarin gescoord wordt. (Bijvoorbeeld: In de categorie 'destructief gedrag' wordt zowel een jongetje ondergebracht dat wil uitproberen of zijn 9-volts radiootje harder speelt op het lichtnet, als het meisje dat bij gebrek aan een krijtje haar naam met een mesje op het schoolbord wil schrijven. Twee verschillende eigenschappen van 'destructief gedrag', die nader benoemd en onderscheiden dienen te worden.) Door deze toekenning van eigenschappen binnen categorieën ontstaan *groepen* (binnen de categorieën).

Vooraf in de beginfase zal men gegevens verkrijgen die niet in een bestaande categorie passen. Dit leidt tot het ontwikkelen van nieuwe categorieën.

Het is vervolgens tijd geworden de categorieën en hun eigenschappen te analyseren. De categorieën worden beter gedefinieerd, evenals de eigenschappen er binnen. Groepen worden binnen de categorieën vastgesteld en de overeenkomsten en verschillen ertussen worden binnen de categorieën en over de categorieën in zo 'objectief' mogelijke *begrippen* beschreven. Uit deze begrippen volgen de eerste hypothesen (gefundeerde veronderstellingen). Deze op hun beurt geven suggesties voor het verzamelen van nieuwe gegevens. De nieuwe gegevens worden weer met de oude gegevens vergeleken; nieuwe categorieën, eigenschappen en groepen ontstaan, bestaande categorieën, eigenschappen en groepen worden weer beter gedefinieerd, nieuwe relaties worden gevonden en bestaande versterkt. Zo ook worden weer nieuwe begrippen beschreven en bestaande begrippen nader geëxpliciteerd. Dit werpt weer nieuwe hypothesen op en verifieert al dan niet de oude hypothesen. Deze verificatie (bevestiging) maakt een hypothese minder hypothetisch en een theorie begint te groeien.

Zo verder werkend – gelijktijdig gegevens verzamelen, analyseren

en theorie ontwikkelen – verkrijgt men een veelzijdig beschreven beeld van het onderzochte fenomeen.

De begrippen, waaruit de hypothesen worden geformuleerd en waaruit de theorie wordt ontwikkeld, moeten voldoende analytisch van karakter zijn (toegankelijk voor systematische beschouwing, het begrip moet ondubbelzinnig zijn). Tegelijkertijd moeten ze een 'invoelbaarheids-aspect' bezitten. Dit laatste kan worden verkregen door met behulp van gepaste illustraties iemand in staat te stellen het begrip in termen van zijn eigen ervaring te vatten. Het doel hiervan is, dat iedereen de theorie en daarmee het onderzochte verschijnsel 'begrijpt' en op basis daarvan in staat is tot beheersing van het verschijnsel (als de mogelijkheden daartoe aanwezig zijn). Een voorbeeld: Als uit een theorie over professionele verzorging valt af te leiden dat de opleiding van verpleegkundig personeel en de structuur binnen de werkring zodanig zijn afgestemd op efficiency, dat daardoor het personeel geen oog en tijd heeft voor (mede)menselijke gevoelens, dan kan deze constatering voor een groot deel van het verplegend personeel een aanzet zijn om te bedenken dat het aspect (mede)menselijkheid ook in hun concrete werksituatie meer aandacht behoeft.

Door het analytische karakter zijn de begrippen toegankelijk voor kritiek en bruikbaar voor verdere algemenering dan alleen het specifieke gebied van het onderzoek. (Het begrip: 'een partijtje spelen' is een algemener begrip in een sportieve krachtmeting dan: 'een stootje maken', wat zich doorgaans tot biljarten beperkt.)

Even terug naar het beginpunt. Misschien kan men zich niet goed voorstellen dat uit doelloos verzamelen van gegevens toch iets samenhangends te voorschijn komt. Barney Glaser en Anselm Strauss, die een boek over deze methode hebben geschreven, benadrukken dat zo'n voorstelling begrijpelijk is: aanvankelijk schijnen hypothesen onderling weinig verband te vertonen. Maar nieuwe categorieën en eigenschappen ontwikkelen zich. Ze krijgen een steeds abstractere betekenis, wat noodzakelijk is om verbanden te kunnen ontdekken en te kunnen interpreteren tussen categorieën en eigenschappen. Uit deze nieuwe verbanden ontstaat een vaster omschreven, meer geïntegreerde structuur, die de theorie buiten het drijfzand houdt. Dat is de essentie van het *zich ontwikkelen* van een gefundeerde theorie. De theorie ontwikkelt zich steeds verder uit de verzamelde gegevens en spoort aan nieuwe gegevens te verzamelen.

Maar wanneer houdt het dan op? Wanneer is het aanvankelijke onderzoeksverschijnsel voldoende beschreven? De methode zelf beslist wanneer het zover is. Er treedt een verzadiging op aan categorieën, doordat er geen nieuwe ontstaan, en het blijkt dat nieuwe gegevens steeds passen in de reeds ontwikkelde categorieën.

Wanneer deze verzadiging zich voordoet, is het van belang dat de onderzoeker(s) nagaat (nagaan) dat andere manieren van gegevensverzameling geen nieuwe categorieën doen ontstaan. Is bijvoorbeeld louter een observatietechniek gebruikt en is het tijdens de analyses van de gegevens niet duidelijk geworden dat ook literatuuronderzoek erbij betrokken moest worden, dan verdient het aanbeveling zich alsnog af te vragen of daarin nieuwe elementen te vinden zijn.

Het gebruiken van verschillende verzamelingstechnieken kan zeer wel nieuwe hypothesen opwerpen, waardoor het onderzochte verschijnsel van een andere zijde wordt belicht. Deze verschillende manieren van tegenaan-kijken worden dan geïntegreerd in de theorie over het verschijnsel.

Tenslotte stelt de funderende methode nog vier opvallende eisen aan de ontwikkelde theorie, die zowel voorwaarde zijn als rechtvaardiging voor het gebruik van deze methode.

1. De theorie moet *werken* op het concrete gebied van het onderzoek. Dit valt doorgaans vrij eenvoudig te controleren: Als de theorie zegt dat treinconducteurs tijdens de avondspits sneller hun geduld verliezen dan op andere tijden van de dag en een eenvoudige steekproef wijst uit dat, zowel tijdens als buiten het spitsuur, 35 van de 50 conducteurs de vriendelijkheid en het geduld zelve zijn bij de gestelde vraag: 'Kunt u mij nu eens even haarfijn uitleggen hoe dit spoorboekje werkt!', dan werkt de theorie niet.
2. De theorie moet *begrijpelijk zijn* voor mensen die ermee moeten werken. Stel de bedrijfsdirectie van een fabriek krijgt van de bedrijfspsycholoog de suggestie bij de expeditie van dat bedrijf een andere werktijdindeling in te voeren. De kans dat dat dan ook gebeurt, zal voor een groot deel afhangen van het feit of de directie begrijpt waarom die werktijdindeling gewijzigd moet worden.
3. De theorie moet *algemeen* genoeg zijn om toepasbaar te zijn in verschillende dagelijkse situaties binnen een ruimer kader dan alleen de specifieke onderzoekssituatie. Als het onderzoek zich bezighield met spelende kinderen in een observatieruimte van de

universiteit tussen 2 en 4 's middags, dan moet de theorie ook geldigheid hebben voor spelende kinderen in het algemeen.

4. De theorie is alleen zinvol als de gebruiker gedeeltelijke *controle* heeft over de dagelijkse situaties waarin de theorie moet worden toegepast. Als de theorie zegt dat schoonmoeders slecht en onveranderbaar zijn, dan hoef je hem niet toe te passen. (De theorie is dan alleen theoretisch interessant, niet praktisch relevant.)

### *Welke methode?*

Hoewel kwantitatief onderzoek een ingewikkelde structuur heeft met al zijn niveaus, restricties en aannamen, is dit eenvoudiger toepasbaar dan de hierboven beschreven methode.

De funderende methode is pas dan funderend als er veel werk in is geïnvesteerd.

Toch is dit niet de reden om sneller te kiezen voor de traditionele wetenschappelijke methode. De keuze wordt bepaald door het verschijnsel dat men wil onderzoeken en door wat men van het verschijnsel wil weten of ermee wil doen.

Weet men niets of weinig van het gebied waarop het onderzoek zich zal richten, dan is het altijd verstandig eerst verkennend onderzoek te doen: men verzamelt gegevens en kijkt ernaar (vergelijken). Het zal blijken of de kenmerken van het verschijnsel zich eerder kwalitatief of kwantitatief laten meten.

Men moet zich afvragen of het te onderzoeken verschijnsel beschreven moet worden, verklaard (kennisvraag) of dat het verschijnsel beheerst moet worden in een praktische situatie.

Kennis van een verschijnsel is een filosofisch probleem en vraagt om voorzichtigheid en controleerbaarheid. Hierbij zal de empirische methode, uit kennis-theoretisch oogpunt, de voorkeur verdienen. Kan er niet gekwantificeerd worden? Jammer! Doet men daarentegen onderzoek dat in het dagelijkse leven toepassing krijgt en dat onmiddellijk probleem-gericht is, dan is het zaak een zoveel mogelijk inzichtelijke theorie te ontwikkelen ten aanzien van het te beheersen verschijnsel.

Controle op de juistheid van de theorie is hier niet een methodologisch probleem, maar een praktijkervaring (de theorie 'werkt' of 'werkt' niet).

Verregaande veralgemenisering van de theorie is meestal niet nodig (soms wel handig!), omdat praktisch onderzoek zich meestal richt op een vrij specifiek gebied.

De funderende methode lijkt met getrainde onderzoekers en de investering van geld, tijd en geduld een betere oplossing te geven voor allerlei praktische maatschappelijke problemen (een oplossing die bovendien begrijpelijk is voor eenieder die kennis neemt van de onderzoeksresultaten), dan de 'gefundeerde wetenschappelijke methode'.

#### *Aanbevolen literatuur*

Delen van de hieronder genoemde literatuur zullen u bekend voorkomen.

B. G. Glaser and A. L. Strauss, *The discovery of Grounded Theory; Strategies for qualitative research*. Chicago, Aldine Publishing Company, 1967.

F. N. Kerlinger, *Foundations of behavioral research*. London, Holt, Rinehart and Winston, 1975.

K. R. Popper, *The logic of scientific discovery*. New York, Harper, 1965.

Meerling, *Methoden en technieken van psychologisch onderzoek*. Meppel, Boom, 1980.

## BESLUIT

We hopen met dit boek te hebben aangetoond dat er veel interessant en belangrijk onderzoek wordt gedaan binnen de sociale wetenschappen, maar ook dat dat onderzoek wordt gedaan door mensen die niet plotseling in een soort computers veranderen op het moment dat zij hun laboratoria binnenschrijven. Ze brengen zichzelf mee, voorzien van een persoonlijkheid, sterke en zwakke punten, overtuigingen en vooroordelen. Al datgene wat die mens onderscheidt van een computer, weerspiegelt zich ook in de resultaten van zijn werk. Die resultaten hebben weer invloed op de maatschappij en op het privé-leven van iedereen. (Denk maar aan de t.v. en de vrijetijdsbesteding, maar ook aan onderwijsvernieuwing en intelligentieonderzoek.)

Iedereen ondervindt de invloed van dit speurwerk en kan, zo lang men onderzoeksresultaten verwacht met de 'waarheid', volop gemanipuleerd worden.

We hopen dat iedereen, na het lezen van dit boek, als iemand hen weer eens probeert te overdonderen met onderzoeksresultaten, onmiddellijk zal vragen: Hoe zat dat onderzoek in elkaar? Wie heeft het onderzoek gedaan? Waar komt hij vandaan? Wat is dat voor figuur? enz. (Deze informatie heeft tot doel het onderzoek beter te kunnen beoordelen, zeker niet om het bij voorbaat te verwerpen (zie inleiding over intelligentie) of te aanvaarden.)

We zijn ons ervan bewust dat wat wij hier behandeld hebben slechts een klein deel is van alle factoren die onderzoeksresultaten beïnvloeden. Met name tijdens de statistische bewerking zijn die vrijwel onbepaald. (Gevraagd naar de reden waarom een onderzoeker een andere bewerking had gekozen dan de volgens de statisticus enig mogelijke, antwoordde deze: 'Omdat mijn resultaten nu veel mooier zijn.')



## REGISTER

- Analyse-eenheden 38, 39, 86  
 antwoordtendentie 75
- Bales, R. F. 36, 46  
 Bartelds, J. F. 69, 78  
 Bend, J. G. van der 143  
 beoordelingsschalen 37  
 betrouwbaarheid 137  
 black box 44  
 Bosch, J. D. 33, 46  
 Bovenkerk, F. 52, 63, 143  
 Brinkgreve, C. 95  
 Brunt, L. 138, 143
- Campbell, D. T. 143  
 Cannell, Ch. F. 63  
 categorieënsystemen 35, 37  
 codeur 89, 90  
 consistentie, interne 107  
 construct 96  
 context-eenheid 88  
 correlatie 22, 23  
 Cottle, T. J. 130  
 criterium-variabele 119, 120  
 Cronbach, L. J. 28, 46
- Diepte-interview 136  
 Drenth, P. J. D. 130
- Edwards, A. L. 112  
 event sampling 38, 39  
 experiment, levensecht 138, 139  
 expert-error 66
- Face-validiteit 93, 120  
 Feyerabend, P. 16, 18  
 Flanders, N. A. 36, 46  
 follow-up 77  
 Fox, D. J. 24  
 Friedrichs, J. 46  
 funderend onderzoek 148-151  
 funnel-approach 68
- Glaser, B. G. 150, 153  
 Gleser, G. C. 85, 87, 88, 90, 92,  
 93, 94, 95
- Gottschalk, L. A. 85, 87, 88, 90,  
 92, 93, 94, 95  
 Groot, A. D. de 18, 24, 26, 46  
 Gutmann-schaal 72, 108-111
- Halo-effect 38  
 Hendriks, A. F. C. J. 46  
 Holsti, O. R. 95  
 Hoogstraten, J. 133, 143  
 Houts, P. L. 130
- Inhoudsanalyse 82, 140, 141, 142  
 analyse-eenheid 86  
 categorieën 87  
 codeurs 89, 90  
 context-eenheid 88  
 gegevensverwerking en  
 interpretatie 93  
 kruisvalidatie 94  
 meeteenheid 86  
 steekproeftrekking 90  
 tel-eenheid 86  
 toepassing 82  
 uitvoering 85  
 validiteit 93  
 voor- en nadelen 83  
 voortest 89
- interactie-analysesysteem van  
 Bales 36  
 interbeoordelaar-  
 betrouwbaarheid 117  
 intercoder agreement 40, 41, 90,  
 117  
 interne consistentie 107  
 invaliditeitsbronnen 123  
 item-test, geprecodeerde
- Joyce, James 84
- Kahn, R. L. 63  
 Kerlinger, F. N. 153  
 Kluiters, H. 78  
 Kok, J. 45  
 Koningsveld, H. 18, 27, 46  
 Korzec, M. 95  
 kritische psychologie 132, 139  
 kruisvalidatie 94  
 Kruijer, G. J. 26, 27, 46
- Leading questions 74

- levensecht experiment 138, 139  
 Likertschaal 72, 106-108  
 Lininger, C. A. 78  
 Lüdtke, H. 46
- Marxisme** 131  
 Meerling 153  
 meer-puntsschalen 100-103  
 meeteenheid 86  
 methode  
   funderende 153  
   gestructureerde 135, 137, 139  
   'harde' 134, 136, 139  
   keuze van 152  
   'vrije' 135, 136, 137, 139  
   'zachte' 134, 139
- Niets-verstorende meting 138
- Observatie** 138, 140, 141, 142  
   gestructureerde 34  
   participerende 113, 136  
   systematische 29, 30, 34, 113  
   vrije 44  
 observatiecategorïeën 34  
 observatiesysteem 34, 35, 39  
   van Flanders 36  
 observeren  
   analyse-eenheden 38, 39  
   beoordelingsschalen 37, 39  
   betrouwbaarheid 39  
   categorïeënschaal 39  
   categorïeënsystemen 35  
   en denken 26  
   direct en indirect 33-34  
   gestructureerd 25  
   gestructureerd en  
     ongestructureerd 30  
   en interpreteren 27  
   openlijk en verborgen 32-33  
   participerend 113, 136  
   participerend en niet-  
     participerend 31-32  
   rating-scales  
     (boordelingsschalen) 37  
   systematisch vs. vrij 43  
   tekensystemen 36, 39  
   validiteit 42  
   en waarnemen 26  
 onderzoek, funderend 148-151
- Paard Hans 50  
 paarsgewijze vergelijking 73, 103-  
   106  
 parallel-test-methode 124  
 Parreren, C. F. van 131, 143  
 pilot study 75  
 Popper, K. R. 27, 153  
 postenquête 76-78  
 psychologie, kritische 132, 139
- Questions, leading 74
- Rating-scales 37  
 relevantie 22  
 respondent orientation 65
- Scalogram** 110  
 schaalconstructie 97  
   aantal schaalpunten 101  
   dimensionaliteit 97-100  
   meer-puntsschalen 100-103  
   omschrijving van  
     schaalpunten 102  
   paarsgewijze vergelijking 103-  
     106  
   schalen voor personen en  
     stimuli 97  
 schaalpunten  
   aantal 101  
   omschrijving 102
- Schwartz, R. D. 143  
 Sechrest, L. 143  
 Segers, J. H. G. 65, 78  
 Selltiz, C. 78  
 significantie 22  
   en relevantie 22  
 Smeden, K. G. van 78  
 sociale wenselijkheid 75  
 spiegelen 62  
 steekproeftrekking 90  
 Strauss, A. L. 150, 153  
 Sumaski, W. 30, 45, 46
- Tanur, J. M. 112  
 Tavecchio, L. W. C. 46  
 tekensystemen 36, 37  
 tel-eenheid 86  
 Terwel, J. 45  
 test 141, 142  
   betrouwbaarheid 118, 122, 123, 124

- constructie 125, 126
- efficiëntie 115, 121
- generaliseerbaarheid 117, 122
- geprecodeerde item-test 125
- invaliditeitsbronnen 123
- normering 115, 122, 125
- norm-groep 116
- objectieve 117
- objectiviteit 116, 122
- parallel-test-methode 124
- relativering 126-130
- reproduceerbaarheid 118
- 'ruis' 114, 126, 127
- standaardisatie 115, 122
- testgedrag 127
- test-her-test-methode 123
- testsituatie 122
- validiteit 119-124
- testgedrag 127
- test-her-test-methode 123
- testsituatie 122
- Thrasher, F. M. 32, 46
- time sampling 38, 39
- trechterechniek 60
- Trevor Roper, H. 81, 83, 95
- triadenvergelijking 73
- Tuckmann, B. W. 78
  
- Unobtrusive measurement 138, 139
  
- Valide 119
- validiteit 42, 137
  - concurrente 42, 93, 120, 121
  - construct 42
  - ecologische 51
  - externe 124
  - face- 93, 120, 121
  - inhouds- 42
  - interne 124, 128
  - predictieve 42, 93, 119
  - voorspellende 42, 93, 119, 121
- Veenman, S. 45
- Ven, A. H. G. S. van der 112
- vergelijking
  - paarsgewijze 73, 103-106
  - triaden- 73
- vooronderzoek 75
- voortest 60, 89
- vraaggesprek 49, 53, 141, 142
  
- afname 61
- directe en indirecte vragen 59, 74-75
- formuleren van vragen 53-57
- motiveren tot deelname 49-50
- open en gesloten vragen 58, 69-74
- registratie 62
- verkrijgen van juiste antwoorden 50-53
- volgorde van vragen 60
- voortest 60
- vragen
  - directe en indirecte 74-75
  - eendimensionele 67
  - eenduidige 67
  - schijnbaar open 70
  - soorten 69
  - suggestieve 67
- vragenlijst 141, 142
  - directe en indirecte 74-75
  - follow-up 77
  - formulieren 65-69
  - open en gesloten 69-74
  - pilot study 75
  - postenquête 76-78
  - schijnbaar open 70
  - soorten 69-75
  - typen van gesloten 70-74
  - vooronderzoek 75
  
- Warwick, D. P. 78
- Webb, E. J. 138, 143
- Wersig, G. 95
- Wrightman, L. S. 78
  
- Zacharias, J. R. 127, 130
- Zeeuw, J. de 130