

# Kwalitatieve analyse van interviews

## Fasen, stappen en computergebruik in intensief onderzoek

F. Wester en V. Peters

### Samenvatting

In dit artikel wordt een beschrijving gegeven van de opzet en uitvoering van een kwalitatieve analyse van interviewmateriaal. Allereerst wordt de kenmerkende cyclus: waarnemen – analyseren – reflecteren, die meer malen wordt doorlopen, toegelicht. Daarna worden de fasen van de kwalitatieve analyse vanuit het perspectief van de gefundeerde theoriebenadering nader uitgewerkt, waarbij steeds de belangrijkste stappen en de knelpunten en valkuilen die tijdens die fase dreigen, worden toegelicht. Daarbij wordt aan de hand van het computerprogramma Kwalitan geïllustreerd op welke wijze de computer kan worden ingeschakeld ter ondersteuning van deze analyse.

### Inleiding

De afgelopen tien jaar is er een sterke ontwikkeling geweest in de beschikbare programma-tuur voor het manipuleren van tekstachtig materiaal. Daarmee is de grootste barrière weggenomen om de computer in kwalitatief onderzoek als hulpmiddel in te schakelen. Dat is zeker het geval voor het kwalitatief onderzoek uit de interpretatieve onderzoekstraditie. Kenmerkend voor dit soort kwalitatief onderzoek is het arbeidsintensieve karakter van de iteratieve, interpreterende analyse van het tekstmateriaal. In deze bijdrage willen wij de systematiek van de kwalitatieve analyse beschrijven en de rol van de computer daarbij toelichten. Eerst zullen wij ingaan op de aard van de kwalitatieve analyse. Vervolgens geven wij een schets van de algemene functies die de computer daarbij kan vervullen. Daarna gaan we specifiek in op de systematiek van waarneming en analyse en de rol van de computer daarbij in onze uitwerking van de *gefundeerde theoriebenadering* (vgl. Wester, 1995). Het gaat hier veelal om vergelijkend onderzoek op grond

van een relatief flink aantal interviews, ook wel aangeduid als kwalitatief survey (Hijmans, 1995, p. 519; Kuyper, 1993; Hijmans, 1994). Ten slotte zullen enkele conclusies getrokken worden over het gebruik van de computer in de kwalitatieve analyse.

### 1 De analyse in het kwalitatieve onderzoek

De analyse<sup>1</sup> is lange tijd het stiefkindje van het kwalitatief onderzoek geweest. Juist op dit gebied zijn er de laatste vijftien jaar veel publicaties verschenen en zijn er recentelijk nieuwe ontwikkelingen, die samenhangen met de rol van de computer in de analyse (Fielding & Lee, 1991; Miles & Huberman, 1993; Weizman & Miles, 1994; Peters & Wester, 1995).

#### 1.1 Kenmerken van de analyse

Voor het kwalitatieve onderzoek zijn twee betekenissen van het begrip analyse relevant. Ten eerste gaat het bij analyse om uiteenleggen, dat wil zeggen dat wordt geëxpliciteerd welke aspecten van het onderzochte verschijnsel relevant zijn. Die aspecten komen naar voren in het analytisch kader van de onderzoeker. In kwalitatief onderzoek moet dat kader veelal aan de hand van het onderzoek worden uitgewerkt. Ten tweede gaat het bij analyse om het zoeken naar patronen in het materiaal, dat wil zeggen de samenhang tussen de onderscheiden aspecten rond een verschijnsel. Via analyse wil men een overzichtsbild krijgen van een verschijnsel en nagaan welke factoren met de aspecten van het verschijnsel samenhangen.

Beide betekenissen van analyse verwijzen naar de rol van het analytisch kader van de onderzoeker. Omdat in kwalitatief onderzoek dit kader veelal nog moet worden uitgewerkt en/of toegespitst op het onderzoeksveld, is analyse sterk met het ontwikkelen van het analy-

tisch kader verbonden. De analyse is dan ook niet één aparte fase in het onderzoek: zij is een steeds terugkerend moment van de spiraalvormige cyclus: reflectie – waarneming – analyse – reflectie, die de onderzoeker tijdens het onderzoek voortdurend doorloopt en waarbij het analytisch kader steeds verder uitgewerkt wordt. In reflectiemomenten wordt nagedacht over onderzoeks- en analysevragen en over het begrippenkader om die te beantwoorden. In de waarneming wordt het materiaal geproduceerd waarin die antwoorden naar voren moeten komen.

De aard van deze handelingen en de ontwikkeling daarin in de loop van het onderzoek brengen met zich mee dat de onderzoeker behoefte heeft aan hulpmiddelen waarin een en ander overzichtelijk kan worden uitgevoerd. Een van de belangrijkste hulpmiddelen voor het reflectieproces zijn de *memo's* waarin de onderzoeker ontwikkelingen vastlegt in begrippen en onderzoeksvragen. Daarnaast zal de onderzoeker het materiaal ordenen en bewerken door te coderen<sup>2</sup>, samen te vatten en overzichten te produceren, waarbij de computer een rol kan spelen.

### 1.2 Analyse en de onderzoeksopzet

Men kan een aantal handelingen onderscheiden, waarvan sommige meer malen in het analyseproces terugkeren.

Een eerste handeling is het *formuleren van de probleemstelling*. In de probleemstelling wordt aangegeven welke aspecten van onderzoeksthema relevant worden geacht en eventueel welke context daarbij een rol speelt. Voor kwalitatief onderzoek betekent dit dat de onderzoeker een beredeneerde vraagstelling heeft uitgewerkt waarin het onderzoekersperspectief veelal in de vorm van *sensitizing concepts* naar voren komt. Dit is dus analysewerk, dat gebaseerd is op een theoretische doordinking van het onderzoeksthema, meestal aan de hand van de onderzoeksliteratuur. In de loop van het onderzoek zal de probleemstelling steeds weer aangepast worden, omdat er ontwikkelingen zijn in het analytisch kader waardoor de onderzoeksvragen explicieter kunnen worden geformuleerd. De probleemstelling is dan ook gedurende het gehele onderzoek een aandachtspunt voor de reflectieprocessen van de onderzoeker. De uiteindelijke probleemstel-

ling kan worden geformuleerd als het analytisch kader is uitgewerkt.

Een tweede handeling is het uitwerken van instrumenten voor de *gegevensverzameling*. Ook hier gaat het om aspecten van het analytisch kader: men wil immers gegevens verzamelen die relevant zijn voor de onderzoeksvragen. Het gaat hier om een aparte stap, omdat via de verkenning van het onderzoeksveld en de onderscheidingen die daarbij relevant zijn, *veldbetrokken begrippen* in het analytisch kader worden opgenomen. Bovendien houdt de gegevensverzameling altijd keuzen in wat betreft de te onderzoeken cases en de kenmerken die daarbij moeten worden vastgesteld. Ook hiervoor geldt dat de onderzoeker in de loop van het onderzoek steeds beter kan formuleren om welke aspecten het gaat en welke gegevens daarvoor verzameld moeten worden.

Een derde activiteit is *transcriptie* van het ruwe gegevensmateriaal. In kwalitatief onderzoek levert de waarneming meestal ruw en voor de onderzoeker ongestructureerd materiaal op, zoals op tape vastgelegde interviews, veldaantekeningen en archiefstukken. Op een of andere manier moet daarvan materiaal gemaakt worden dat geschikt is voor analyse. Bij transcriptie spelen weer keuzen een rol die consequenties hebben voor de mogelijkheden in de analyse (bijvoorbeeld nemen we naast de gesproken woorden ook aarzelingen en houdingen mee). Vooral bij observatiemateriaal zijn expliciete regels nodig omtrent het wat en hoe van de registratie van de geobserveerde verschijnselen.

Aan de transcriptie is tevens een activiteit verbonden die meer technisch van aard is: een *eerste ordening* van het materiaal, zodat het geschikt is voor de analyse-in-engere-zin. Tegenwoordig gaat het hier meestal om een zodanige bewerking van het ruwe materiaal, dat we de computer als hulpmiddel kunnen inschakelen. Zo vragen sommige programma's voor kwalitatieve analyse dat men het materiaal bijvoorbeeld indeelt in segmenten en zonder tekstverwerkingscodes bewaart.

Een vierde analysehandeling is dan *de analyse-in-engere-zin*. Globaal genomen gaat het hier om het lezen van een sample van het onderzoeksmateriaal. Hier zijn verschillende betekenissen van *lezen* van belang. Ten eerste gaat het natuurlijk om lezen in de zin van het

waarnemen van verbaal weergegeven betekenissen (zoals we een krant of boek lezen). Daarnaast gaat het om selectief lezen vanuit de onderzoeksvragen (c.f. aren lezen, uitgelezen waar). Bovendien gaat het om interpreterend lezen, dat wil zeggen dat we interpreteren vanuit het analytisch kader (c.f. kaartlezen, sterren lezen). Deze lezingen van het materiaal worden meestal uitgedrukt in codes of trefwoorden die bij het materiaal worden opgenomen. Via reflectie op vragen, materiaal en codes worden codes aangepast en nieuwe analysevragen geformuleerd. Deze worden dan weer op een volgende selectie uit het materiaal uitgeprobeerd, waarbij een nieuwe cyclus van materiaalselectie, lezen, interpreteren en reflecteren is ingezet. Dit cyclische proces gaat door tot de onderdelen van het analytisch kader zijn uitgewerkt.

Een vijfde analysehandeling zijn de *nadere analyses* die meestal gericht zijn op het vergelijken van cases via overzichten, tabellen en matrices om patronen in het materiaal te ontdekken.

De laatste analysehandeling die wij onderscheiden is de *rapportage*, een aandachtspunt in elke onderzoeksfase. Via tussentijdse verslagen kan de voortgang van het onderzoek worden bewaakt. Het schrijven van het eindverslag, waarin de beantwoording van de definitieve onderzoeksvraag wordt beargumenteerd, vormt de afronding van de analyse.

Door analyse op te vatten als een steeds terugkerende moment in een cyclisch proces wordt een aantal kenmerken van de kwalitatieve analyse duidelijk. Ten eerste wordt de analyse altijd gestuurd door de *vragen* die in het reflectieproces naar voren zijn gekomen en de observaties hebben gestuurd. In het verzamelde materiaal worden antwoorden op die vragen geformuleerd door middel van het voorlopige *analytisch kader*, dat daartoe eventueel wordt aangepast, en waarmee het materiaal wordt geordend en bewerkt. Ten tweede is de analyse in de loop van het onderzoek niet steeds op dezelfde vragen gericht. De spiraalvormige cyclus duidt een ontwikkeling in het onderzoek aan, waarbij de opgedane inzichten een nieuwe cyclus aansturen, waarbij vragen veranderen. Bovendien kan men de loop van het onderzoek uiteenleggen in een aantal fasen waarin achter-

envolgens een aantal subdoelen worden bereikt. Bij elk subdoel staat een ander soort vragen centraal.

Zo kan men in een onderzoek volgens de gefundeerde-theoriebenadering (Glaser & Strauss, 1967; Strauss & Corbin, 1990) een aantal fasen in het theorieontwikkelingsproces onderscheiden. In onze uitwerking van de gefundeerde-theoriebenadering worden de volgende fasen met doelstellingen onderscheiden:

- exploratiefase: de onderzoeker formuleert een probleemstelling en probeert aan de hand van een verkenning van het veld zo veel mogelijk relevante begrippen te formuleren;
- specificatiefase: de meest centrale begrippen worden uitgewerkt in termen van dimensies en variabelen waarmee het veld kan worden beschreven;
- reductiefase: de kern van de theorie en de relatie ervan met de centrale begrippen wordt geformuleerd;
- integratiefase: de theorie, als antwoord op de definitieve probleemstelling, wordt systematisch onderzocht.

In deze vorm van kwalitatief onderzoek staat de uitwerking van het analytisch kader centraal. De fasen kunnen als stappen in dat proces worden opgevat. Dit leidt ertoe dat de analyse in elke fase op de beantwoording van andersoortige vragen is gericht.

Deze in globale termen beschreven analysehandelingen komen min of meer terug in analyseprocedures die door verschillende auteurs in specifiekere stappen zijn beschreven (bijv. Spradley, 1979, 1980; Miles & Huberman, 1984, 1993; Maso, 1987; Wester, 1987, 1995). Recent zijn daar weer nadere uitwerkingen bij gekomen, mede onder impuls van de ontwikkeling van een groot aantal computerprogramma's ten behoeve van kwalitatief onderzoek. Men kan hierbij denken aan programma's als Ask Sam, The Ethnograph, Nudist, Atlas-ti en

Kwalitan (voor een consumentenbondachtige beschrijving zie Weizman & Miles, 1994; zie ook Muhr, 1992; Richards & Richards, 1990 en Seidel, 1988).

In elke fase van het analyseproces kan de computer worden ingeschakeld om de onderzoeker te voorzien van relevant materiaal, om analysebewerkingen te verwerken en de bevindingen vast te leggen. De analysestappen kunnen daardoor verder worden gespecificeerd (voor een uitwerking binnen de gefundeerde theoriebenadering zie Wester, 1995). Daarbij is de computer een zeer belangrijk hulpmiddel, die regelmatig wordt ingeschakeld bij allerlei analyseactiviteiten. Maar het is belangrijk te beseffen dat de computer een hulpmiddel blijft: de feitelijke analyse wordt door de onderzoeker zelf uitgevoerd.

## 2 Nadere uitwerking van de analyse volgens de gefundeerde theoriebenadering

Na de beschrijving van de terugkerende handelingen en de fasering van de analyse gaan wij hier specifiek in op de verschillende fasen van de gefundeerde theoriebenadering. In feite gaan we concreter in op hetgeen in de vorige paragraaf werd aangeduid als de analyse-inengere-zin. Waar het gaat om de ondersteuning van de computer refereren wij daarbij aan het programma Kwalitan (Peters, 1995), dat speciaal voor dit soort analyse is ontwikkeld. Men zou dezelfde ondersteuning in principe ook via andere programma's kunnen realiseren, hoewel er tussen de diverse programma's grote verschillen bestaan in de manier waarop het materiaal geprepareerd moet worden, de mogelijkheden die het programma biedt en de wijze waarop men met het programma communiceert.

In de volgende uitwerking volgen we de eerder omschreven fasen in het onderzoeksproces, waarbij we beginnen met bewerkingen op het materiaal nadat een eerste hoeveelheid materiaal is verzameld. We gaan er daarbij vanuit dat het gaat om een kwalitatief survey waarbij de gegevens verzameld worden via een flink aantal interviews. We zullen dan ook aangeven hoe de interviewstrategie er in de verschillende fasen uitziet.

### 2.1 Voorbereiding

Het geschikt maken van het materiaal voor een kwalitatieve analyse houdt allereerst in dat gezorgd moet worden voor de invoer van het materiaal in een computerbestand. Dat betekent meer dan alleen maar het uittypen of scannen van een tekst. Het gaat hier om wat eerder de transcriptie van het ruwe materiaal is genoemd. De onderzoeker moet bepalen hoe het materiaal in een bestand wordt vastgelegd. Dat lijkt voor interviews nogal simpel, maar in feite moet de onderzoeker allerlei beslissingen nemen, zoals het wel of niet opnemen van aarzelingen, stiltes en dergelijke in het transcript, het corrigeren van versprekingen, het opnemen van stemverhogingen, of juist het globaal uitschrijven wat zoal bedoeld is te zeggen. Als het gaat om video-opnames zijn de beslissingen die de onderzoeker moet nemen al veel complexer, omdat visuele informatie in een tekst moet worden omgezet. Maar zelfs als het ruwe materiaal teksten betreft, zoals de notulen van vergaderingen of een beleidsrapport, is er de vraag of men de gehele tekst wil opnemen in de transcriptie of dat men zich beperkt tot de relevante gedeelten daaruit.

Daarnaast delen wij het ruwe materiaal op in segmenten. We zien dat bij vrijwel alle computerprogramma's in de pre-analytische fase een of andere vorm van voorstructurering noodzakelijk is. In het geval van Kwalitan betekent dat, dat het materiaal vooraf wordt ingedeeld in segmenten; deze segmenten kan het programma vervolgens als eenheden onderscheiden. Om te voorkomen dat de onderzoeker van het onderscheiden van segmenten als het ware een aparte analysefase gaat maken, adviseren wij de tekst in segmenten uiteen te leggen aan de hand van 'logische posities' in de tekst, bijvoorbeeld als een interviewer een nieuwe vraag stelt, of in een document als het gaat om een nieuwe alinea of paragraaf. Dit onderscheiden van tekstsegmenten maakt het mogelijk enige structurering aan het materiaal te geven, die de start van de interpretatieve analyse bevordert, bijvoorbeeld omdat het mogelijk maakt dat segmenten over hetzelfde onderwerp met elkaar worden vergeleken. In programma's als Kwalitan is het altijd mogelijk de splitsing in segmenten aan te passen indien de analyse dat nodig maakt (bijvoorbeeld omdat je later meer over het onderwerp weet).

Indien de onderzoeker in dit stadium van de analyse behoefte heeft aan een nadere verkenning van het materiaal, kan hij daarvoor ook weer een beroep doen op de computer. Zo heeft Kwalitan een optie, om een overzicht te maken van de woorden die voorkomen in de tekst en de frequenties waarin ze voorkomen. Hoewel dit een analyseprocedure is die meer past bij vormen van inhoudsanalyse en als zodanig weinig verbindingen heeft met de gefundeerde theoriebenadering, kan een dergelijk overzicht toch nuttig zijn. Op deze manier kan de onderzoeker bekend raken met de soorten woorden en uitdrukkingen die respondenten gebruiken of die in de documenten voorkomen. Inspectie van deze lijst zet de onderzoeker op het spoor van relevante veldbetrokken begrippen.

## 2.2 De exploratiefase

De belangrijkste analyseactiviteit van de onderzoeker in deze fase wordt *open coderen* genoemd. Hierbij worden de teksten gelezen en geïnterpreteerd, worden er (nieuwe) codes geformuleerd en aan de segmenten toegevoegd.

Er kunnen hier verschillende technieken worden toegepast bij het open coderen. De belangrijkste is *gericht lezen*. Daartoe recapituleert de onderzoeker welke begrippen en categorieën bij de uitwerking van de probleem-

stelling, de selectie van respondenten en de uitwerking van de topiclijst een rol hebben gespeeld. Met dit voorlopig analytisch kader wordt het materiaal gelezen, vanuit de vraag: wat wordt er over het betreffende thema (begrip, categorie) gezegd. Deze begrippen en categorieën worden aan de interviewtekst verbonden waar dat relevant is en eventueel worden specifiekere codes geformuleerd die aangeven hoe erover het betreffende thema wordt gesproken en wat erover wordt gezegd. Deze manier van lezen sluit dus sterk aan bij het perspectief van de onderzoeker.

Een tweede techniek is *samenvattend lezen*. Daartoe leest de onderzoeker elk fragment en probeert met een of meer codes de strekking van de tekst en de belangrijkste termen die erin voorkomen aan te geven. Deze codes worden aan de fragmenten toegevoegd. Deze manier van lezen sluit meer aan bij het perspectief van de respondent.

Een derde techniek is *materiaalscanning*. Hier maken we expliciet gebruik van de mogelijkheden van de computer als tekstverwerker om een overzicht te krijgen van het woordgebruik van de respondent. De meeste woorden zijn uiteraard niet specifiek relevant voor het centrale onderzoeksthema, maar van bepaalde woorden is dat wel het geval. Deze woorden kunnen in de overzichtslijst worden geselecteerd.

Werkbestand	Segment	Codes	Output	Memo	Inhoud	Bestand	Overig
			----- Toon segmenten -----				
			+----- tekst van het segment -----+				
+---- huidig segment ----+			I:Leer jij voor goede cijfers?				
documentnr. : 2			Hi:Ja.				
documentnaam : lint119			I:Kan je zeggen waarom je dat doet?				
segmentnummer: 9			Ja, ik vind het niet zo leuk om naar school te gaan eigenlijk, dan denk ik van ja, toch heb je die ondergrond eigenlijk nodig voor je latere opleiding.				
+---- bewerk de codes ----+			Des te hogere cijfers je hebt des te eerder word je aangenomen.				
1. lesstof			Dus waarom ga je niet leren om die cijfers te halen en dan een beetje mooi rapport te krijgen?				
2. cijferleren			Ja en ook om niet te blijven zitten en nog langer op die school te blijven.				
3. kritische acceptatie			Maar ik ga echt niet leren van leuk om te weten of zo, en ik hoef het eigenlijk niet te leren.				
4.			Ik denk dat de meeste mensen eigenlijk leren om de repetities en om een goed cijfer te				
5.							
6.							
7.							
8.							
9.							
10.							
11.							
12.							
			-----				
werkbestand konver #doc 4 #seg 41							

Figuur 1. Code-invoerscherm met segment uit een voorbeeld interview

teerd en als codes automatisch aan de fragmenten waarin ze voorkomen worden gekoppeld. Deze fragmenten kunnen vervolgens worden geselecteerd voor nadere lezing in gerichte en samenvattende zin.

De codes die tijdens het open coderen worden geformuleerd (zie de drie manieren van lezen hiervoor), worden gekoppeld aan de segmenten waarin het thema dat met de code wordt aangeduid, voorkomt. Figuur 1 laat zien hoe een en een ander er bij Kwalitan uitziet. Op het scherm is de informatie met betrekking tot één segment weergegeven. Linksboven staat aangegeven welk segment momenteel wordt getoond. Rechts op het scherm staat de tekst geprojecteerd, terwijl aan de linkerkant ruimte is om de codes in te voeren. Aan elk segment kan een groot aantal (25) codes worden verbonden.

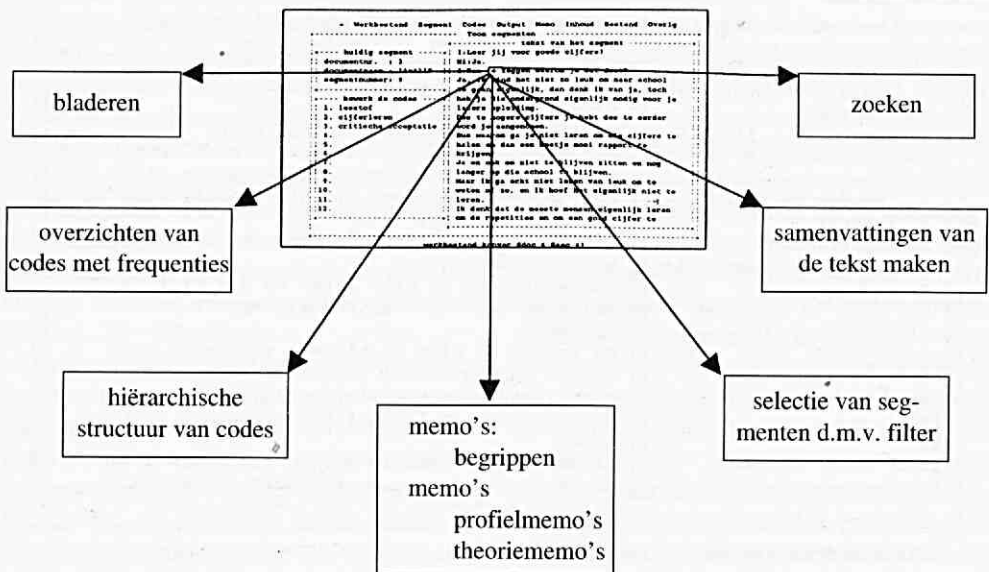
In Figuur 2 staan de verschillende mogelijkheden aangegeven waarmee vanuit het basis-scherm voor het coderen van het materiaal naar andere mogelijkheden van de computer kan worden overgegaan. Zo zijn er enkele opties om door de tekst te bladeren, om samenvattingen te maken van de tekst en de segmenten binnen het werkbestand achtereenvolgens door te lopen (bijvoorbeeld om codes te controleren die men eerder heeft gebruikt).

Eerder is al gewezen op de mogelijkheden

om activiteiten en beslissingen bij te houden in memo's<sup>3</sup>. In de exploratiefase zullen vooral begrippenmemo's worden geproduceerd, waarin de betekenis van de (belangrijkste) codes worden besproken. Daarnaast moeten in theoretische memo's de relaties tussen de codes met het analytisch kader worden bijgehouden. Kwalitan biedt de mogelijkheid om memo's in te voeren, aan te passen en tijdens de analyse op te roepen.

Nadat de codes aan de segmenten zijn toegevoegd, zal de onderzoeker graag een overzicht van de toegekende codes willen hebben. Via de computer vragen we een alfabetisch geordende lijst op van codes die aan de segmenten zijn toegevoegd (zie Figuur 3). Deze lijst geeft de onderzoeker de mogelijkheid te overzien welke codes hij gebruikt om het materiaal te beschrijven, met andere woorden de lijst laat de bouwstenen zien voor het analytisch kader. Daarnaast stelt deze lijst de onderzoeker in staat om typfouten in de codes te corrigeren en na te gaan of er mogelijk synoniemen voor hetzelfde thema zijn gebruikt.

De verkennende fase levert aldus een groot aantal codes op, die voor een deel met het voorlopige analysekader van de onderzoeker zijn verbonden. Daardoor krijgt de onderzoeker een beeld van de centrale begrippen of hoofdthe-



Figuur 2. Code-invoerscherm met segment uit een voorbeeld interview

= Werkbestand	Segment	Codes	Output	Memo	Inhoud	Bestand	Overig
-----				Overzicht	codes	-----	
							1
	acceptatie		1				
	actuele relevantie		9				
	beh. vaardigh. leren		1		bereik : alle documenten		
	behoefte act. relev.		2		volgorde : alfabet		
	cijferleren		9				
	critische acceptatie		2		aantal verschillende codes:		25
	groeps vrijheid		4		totaal aantal codes		97
	groepsafhankelijkh.		2		aantal segmenten		41
	groepsefficiëntie		2				
	groepssamenwerken		3		gemiddelde frequentie		3.88
	huiswerk		1		codes per segment		2.37
	individueel ideaal		1				
	ineff.tijdsstructuur		1				
	inter. vrijw. leren		1				
	interesse		3				
	irrelevantie		1				
	klassikaal saai		2				
	leerkr. ontsp. orde		4				
	leerkracht gelijke		1				
-----							
werkbestand konver #doc 4 #seg 41							

Figuur 3. Overzicht van codes in het werkbestand

ma's op het veld van onderzoek. De verkennende fase wordt afgesloten wanneer nieuw materiaal of een nieuwe lezing van het materiaal geen nieuwe hoofdthema's voor analyse meer opleveren.

Bij deze globale beschrijving van de werkwijze in de verkennende fase kan een aantal aantekeningen worden gemaakt, gelet op de problemen die onderzoekers zoal tegenkomen.

#### a. Het 'tijdverlies' door het transcriptie werk

Het transcriberen van interviews vraagt meer tijd dan onderzoekers vaak gepland hebben en levert een grote papierberg op, waardoor ook de analyse wordt bemoeilijkt. Dat roept de vraag op of daar niet op bezuinigd kan worden. Dat is zeker het geval voor de integratiefase van het onderzoek, maar niet in de beginfasen. Een eerste transcriptie van tape naar flop kan eventueel door hulp personeel worden uitgevoerd, maar de onderzoeker moet zelf de tape afluisteren om het basistranscript te corrigeren. Dat is geen verloren tijd, maar een eerste stap om thuis te raken in het materiaal. Daarbij komt dat de tekst eventueel kan worden aangevuld met toelichtingen uit de context van het interview om onbegrijpelijke of verwarrende teksten te verduidelijken. Na dit correctiewerk is de onderzoeker weer helemaal in het vraaggesprek en is dus in staat een profielmemo te

maken waarin in globale lijnen de plaats van de respondent bij belangrijke topics staat weergegeven.

#### b. Het tijdverslindende exploratieve lezen

De analyse volgens eerder genoemde technieken neemt zeer veel tijd in beslag. Wanneer in het onderzoek zo'n vijftig vraaggesprekken zijn gepland die alle op die manier gelezen moeten worden, zou het tijdbudget van het onderzoek makkelijk overschreden kunnen worden. Daarbij komt dat dit zoveel gecodeerd materiaal oplevert, dat ook de specificatiefase erg veel tijd vraagt.

De oplossing is dat men de veldverkenning op een beperkt aantal strategisch gekozen interviews uitvoert. Dit veronderstelt dat men in de veldverkenning weliswaar veel gesprekken heeft gevoerd (zeg 15), maar dat men aan de hand van de transcripties en de geproduceerde profielmemo's een zodanige selectie maakt (zeg 5) dat er een grote variatie is in wat naar voren is gebracht wat betreft de centrale onderzoekstopics. Immers in de veldverkenning gaat het erom zoveel mogelijk categorieën te ontwikkelen.

#### c. Het gevaar van fragmentatie

De line-by-line analyse en vooral het toekennen van codes aan fragmenten en de mogelijk-

heden die de computer biedt om fragmenten te vergelijken dwars door het materiaal heen, hebben het gevaar in zich dat de teksten worden gelezen los van hun context: het interview met een bepaalde respondent. De oplossing is dat men voortdurend de *profielmemo* van de respondent raadpleegt en dat deze profielmemo in iedere analyseronde wordt aangevuld wat betreft de score van de respondent op de centrale topics.

#### **d. De sturing van de analyse**

Het belangrijkste probleem is dat de onderzoeker leest en leest en steeds nieuw materiaal verzamelt en daardoor ook steeds nieuwe codes kan formuleren en dat oude codes weer steeds verder gespecificeerd kunnen worden, zodat het proces eindeloos voort kan gaan. Het verzadigingsprincipe (er komen geen nieuwe codes bij) werkt niet vanzelf. De onderzoeker moet het analyseproces sturen door naast interviewen en analyseren regelmatig tijd te reserveren om te reflecteren op de bevindingen. Hier kan de onderzoeker gebruik maken van een overzicht van de gebruikte codes en memo's schrijven over de relatie met het analytisch kader. De belangrijkste aanwijzing is Strauss' adagium (Strauss, 1987) 'think theoretical', waardoor structuur in het analytisch kader kan worden gebracht. De veldverkenkende fase kan worden afgesloten als de onderzoeker een voorlopig globaal antwoord op de probleemstelling kan formuleren en dus een idee heeft welke onderwerpen, begrippen en categorieën centraal staan in het onderzoek.

### **2.3 De specificatiefase**

De specificatiefase is erop gericht de begrippen en codes die geformuleerd zijn rond de centrale onderwerpen in het onderzoek, precies uit te werken. De analyse is dus gericht op het ordenen van het analytisch kader om een beschrijving te kunnen geven van de aspecten die rond de centrale onderwerpen in het onderzoek van belang zijn.

De selectie van respondenten is hier gericht op aanvullingen van de onderzoeksgroep die in de veldverkenkende fase is ondervraagd. Het kan hier dus gaan om cases die juist verschillen van bepaalde onderzochte gevallen, maar ook om meer van dezelfde, die dus op andere cases lijken. De zo samengestelde onderzoeksgroep

maakt dan vergelijkingen mogelijk binnen een subgroep van gelijken én tussen subgroepen die verschillen.

Net als in de veldverkenkende fase zal de interviewer de topics bespreken door de respondent te laten uitleggen en toelichten, maar nu heeft de interviewer het voordeel dat hij/zij al wat thuis is in (de variaties in) het idioom waarmee over de onderwerpen wordt gepraat. Daarbij komt dat de analyse tot dan toe een globaal beeld heeft opgeleverd van centrale onderwerpen en bij sommige onderwerpen misschien ook al welke variaties daarbij een rol kunnen spelen. De analyse in de veldverkenkende fase *bevordert zo de mogelijkheden om door te vragen*. De interviewer is er nu op uit bij elk gespreksonderwerp een samenvatting te maken van het perspectief van de respondent, waarbij zoveel mogelijk al termen worden gebruik uit het analytisch kader, dat nu immers ook veldbetrokken begrippen bevat. Daarmee wordt tijdens de veldwerkfase al getoetst of de ontwikkelde begrippen passen op het veld van onderzoek.

Ook van deze vraaggesprekken wordt weer een volledig transcript gemaakt, dat in de computer kan worden ingelezen. De fragmenten worden nu onderscheiden aan de hand van de onderwerpen die in de veldverkenkende fase centraal zijn komen te staan in het onderzoek. Elk fragment krijgt aldus een of meer hoofdcodes, die op de betreffende onderwerpen slaan.

De lezing/codering van het materiaal geschiedt nu vanuit de overzichtslijst van codes, die in de veldverkenkende fase zijn gebruikt. Nieuwe codes worden alleen dan geformuleerd, wanneer aan de overzichtslijst (zie Figuur 3) geen passende code kan worden ontleend.

Is elk interview aldus gelezen en gecodeerd, dan gaat de onderzoeker over tot *vergelijkende analyses*. Hierbij worden alle interviews gebruikt die tot dan toe zijn gecodeerd.

Het belangrijkste doel van de specificatiefase is het uitwerken van centrale begrippen waarmee het verschijnsel dat men onderzoekt kan worden beschreven en de vraagstelling kan worden beantwoord. De analysehandelingen van de onderzoeker concentreren zich op voortdurende vergelijkingen binnen het materiaal. Op basis van de resultaten van de exploratiefase en van de reflecties op die resultaten



(memo's) kiest de onderzoeker een beperkt aantal hoofdcodes, die belangrijk lijken te zijn en mogelijk een centraal begrip in het analytisch kader kunnen worden. Voor elke geselecteerde hoofdcode wordt (als het ware) een bestand gemaakt van de segmenten waar die code aan toe is gevoegd. Dit maakt het mogelijk om te zoeken naar verschillen en overeenkomsten tussen deze segmenten, die door de overige codes bij elk segment worden geïndiceerd. Op basis van deze vergelijkingen kan het begrip dat door de hoofdcode wordt aangegeven nader worden uitgewerkt.

In deze fase van de analyse bestaat de ondersteuning door de computer uit de mogelijkheid om criteria te specificeren om segmenten te selecteren voor de vergelijkingen via een filterprocedure: die fragmenten tekst worden gezocht, die voldoen aan kenmerken die door de onderzoeker worden opgegeven.

Zoals in Figuur 4 wordt weergegeven, zijn er diverse ingangen om criteria te specificeren waarmee de gewenste segmenten uit het onderzoeksmateriaal kunnen worden geselecteerd.

De belangrijkste ingang van het filter is de lijst van een of meer codes. De computer geeft alleen die segmenten weer op beeldscherm (of geprint) die aan de specificaties in het filter voldoen. Als je twee of meer codes in een filter opgeeft kunnen alle combinaties gemaakt wor-

den volgens de logische operatoren 'en', 'of' en 'niet'. Andere specificatieingangen in het filter zijn 'tekststukjes' (woorddeel) die in de tekst van het segment moeten voorkomen, het nummer van het segment en het documentnummer. Ook (zie in de rechter kolom van het venster in Figuur 4) kan informatie opgegeven worden die voorkomt in de documentinformatieregel om de gewenste segmenten te selecteren; als in die informatieregel bijvoorbeeld het geslacht van de respondent is aangegeven, kunnen via deze rubriek de vrouwelijke respondenten worden geselecteerd.

In het voorbeeld zijn alleen die segmenten opgevraagd, die zowel 'acceptatie' als 'toekomstig relevant' als code hebben (naast eventuele andere codes). De andere criteria zijn hier verder niet gespecificeerd.

Het selecteren van segmenten stelt de onderzoeker in staat om verschillen en overeenkomsten tussen de segmenten te thematiseren en aldus het betreffende centrale begrip uit te werken.

Reflectie op het geheel van nevenscodes bij elk centraal begrip brengt de onderzoeker tot formulering van de *dimensies* van het centrale begrip waar de codes naar verwijzen.

Een handig hulpmiddel hierbij is een overzicht van de hiërarchische structuur van hoofd-

```

≡ Werkbestand Segment Codes Output Memo Inhoud Bestand Overig
+----- Filter en toon segmenten -----+
+----- specificieer de opbouw van het filter -----+
| codes | woorden in de tekst |
| 1. acceptatie | ---> |
| 2. toekomstig relevant | |
| 3. | documentnummers |
| 5. | |
| 6. | segmentnummers |
| 7. | ---> |
| 8. | |
| 9. | gegevens uit de documentidentificatie |
| 10. | ---> |
+-----+
| vul maximaal 10 codes in |
+-----+
werkbestand konver #doc 4 #seg 41

```

Figuur 4. Het maken van een filter

```

= Werkbestand Segment Codes Output Memo Inhoud Bestand Overig
+----- Boomstructuur codes -----+
| boomstructuur c:\11_persp           alle niveaus getoond 2 / 23 |
+-----+
> lesstof
  * relevantie
    * korte termijn
      + actuele relevantie
      + behoefte act. relev.
      + interesse
    * middellange termijn
      + cijferleren
      + diplomaleren
    * lange termijn
      + toekomstig relevant
      + nuttig voor beroep
-----
> lesvorm
  * groepen
    * groepsefficiëntie
    * groepssamenwerken
    * leren>gezelligheid
-----
werkbestand konver #doc 4 #seg 41

```

Figuur 5. De hiërarchische structuur van codes

en nevenscodes (zie Figuur 5). In zo'n overzicht zou eventueel tussen hoofd- en nevenscodes de dimensie als tussenlaag kunnen worden gedefinieerd. In het voorbeeld wordt deze tussenlaag gevormd door codes als 'korte termijn', 'middellange termijn' en 'lange termijn'. Deze codes kunnen vervolgens weer gebruikt worden om alle segmenten te selecteren die onder de codes van de 'tussenlaag' vallen.

Deze dimensies worden gebruikt om de *variabelen* te formuleren waarmee verschillen en overeenkomsten tussen respondenten met betrekking tot betreffende centrale onderwerp worden uitgedrukt. Daartoe moet de analyse vanuit deze dimensie per respondent worden uitgevoerd. Voor elke respondent wordt een profielmemo over de dimensie geschreven; vergelijking tussen de respondenten moet een variabele opleveren.

Zoals altijd moeten de reflecties die bij het uitwerken van begrippen en dimensies een rol hebben gespeeld in memo's worden vastgelegd.

Zijn voor alle centrale onderwerpen rond de probleemstelling dergelijke variabelen uitgewerkt, dan is de specificatiefase afgerond. Het analytisch kader is nu voldoende uitgewerkt om een beschrijvend overzicht te geven van de meest centrale onderwerpen in het onderzoek.

Bij deze globale beschrijving van de werk-

wijze in de specificatiefase kunnen de volgende aantekeningen worden gemaakt gelet op de problemen die de onderzoeker kan tegenkomen.

#### a. *Onvolledige interviews*

Na de specificatiefase is de onderzoeker goed in staat aan te geven welke onderwerpen in het interview aan de orde moeten komen en welke begrippen en categorieën daarbij een rol spelen. Dit heeft als voordeel dat het hierdoor mogelijk wordt om interviewers precies te instrueren zodat men (een deel van) de verdere vraaggesprekken kan uitbesteden. Een tweede consequentie van het uitgewerkte analytisch kader is dat de onderzoeker geconfronteerd wordt met lacunes in de vraaggesprekken die tot nu toe zijn gevoerd. Dat hangt samen met de start van het onderzoek: hoe opener het onderzoek, hoe meer lacunes in het analytisch kader, hoe groter de kans dat onderwerpen niet of niet volledig of niet diepgaand genoeg in de interviews aan bod zijn gekomen, gelet op de *uit-eindelijke analyse*. Met sommige respondenten zullen dus aanvullende gesprekken moeten worden gevoerd in de integratiefase.

#### b. *Loskomen van de codes*

De specificatiefase is erop gericht het perspectief van de onderzoeker verder uit te werken.

Dit betekent dat niet alle codes uit de verkennende fasen even belangrijk zijn. Sterker nog, het gaat niet om de codes op zich, ze zijn slechts *een hulpmiddel* om het begrippenkader van de onderzoeker uit te werken. Het analytisch kader bevat na de specificatiefase termen (variabelen, categorieën) die maar voor een klein deel met eerder gebruikte codes samenvallen. Daarbij komt dat in de verkennende fase codes zijn gebruikt, die later niet meer terugkeren omdat het niet om centrale onderwerpen in het onderzoek gaat. De onderzoeker moet dus loskomen van de codes en moet voorkomen dat de analyse uitloopt in een bureaucratisch afwerken van alle losse draden die met de codes zijn gegeven.

#### **c. Veldbetrokken literatuur**

Bij het uitwerken van het analytisch kader heeft de onderzoeker behoefte aan een bredere oriëntering dan het materiaal alleen. Bij de start van het onderzoek heeft de onderzoeker de probleemstelling gebaseerd op een globaal overzicht van het onderzoeksveld op grond van overzichtsartikelen. In de specificatiefase heeft de onderzoeker echter behoefte aan *specifieke literatuur met betrekking tot het onderzoeksveld*, waardoor de relevantie van de veldbetrokken begrippen ook buiten het eigen onderzoeksmateriaal kan worden onderzocht of eventuele onduidelijkheden in begripsformuleringen kunnen worden opgelost. Ook kunnen aan die literatuur begrippen ontleend worden om relevante onderscheidingen in het materiaal te benoemen.

#### **d. Aandacht voor reflectie**

In de specificatiefase heeft de onderzoeker de meeste voordelen van het werken met de computer. In de vergelijkende analyse werkt de onderzoeker met selecties uit het materiaal, die vanuit de analytische categorieën worden vergeleken (cf. Strauss & Corbin, 1990: "axial coding"). Bij die lezing van de fragmenten gaat het minder om het samenvatten en meer om het interpreteren van de tekst. Dit betekent dat het handwerk wordt afgewisseld door denkwerk waarbij de inzichten en de creativiteit van de onderzoeker een grote rol spelen. Reflectie op het analytisch kader en vastleggen van de zich ontwikkelende inzichten in memo's vormen een belangrijk deel van de analyseactiviteiten in de specificatiefase.

## **2.4 De reductiefase**

In de reductiefase wordt de kern van de te ontwikkelen theorie bepaald om structuur te kunnen aanbrengen in de ontwikkelde variabelen en hun samenhangen. In de specificatiefase is het analytisch kader 'verticaal' ontwikkeld dat wil zeggen dat voor elk onderwerp in het onderzoek de structuur van begrippen, dimensies, variabelen en categorieën die daarop betrekking hebben, is uitgewerkt. In de reductiefase gaat het om een 'horizontale' uitwerking, dat wil zeggen de samenhang tussen de onderwerpen met elkaar en dus de manier waarop de begrippen en variabelen met elkaar zijn verbonden.

Wat betreft de kern van de theorie kan het gaan om een centraal begrip, bijvoorbeeld een typologie die de onderzoeker heeft ontwikkeld, die met vele andere variabelen samenhang vertoont. Ook kan het gaan om een proces, dat een ontwikkeling en samenhang beschrijft waarmee de verschillende variabelen kunnen worden geordend. Hier wordt weliswaar gesproken van een aparte fase om het kernbegrip te formuleren, maar vaak is de kern van de theorie in globale zin al tijdens de voorgaande fasen duidelijk geworden. De analyse in deze fase is dan alleen nog maar een toets of het betreffende kernbegrip inderdaad de samenhang in het materiaal brengt.

Ook in de reductiefase kunnen eventueel aanvullende interviews gedaan worden. De selectie van respondenten is hier gericht op het zoeken naar gevallen waarin de samenhang tussen de onderzochte verschijnselen naar voren moet komen. Het zal duidelijk zijn dat het uitgewerkte kader betekent dat de interviewer al veel meer systematisch te werk kan gaan en bij elk topic uit de vragenlijst door zal vragen vanuit de ontwikkelde begrippen en categorieën.

De transcriptie van de vraaggesprekken is hier ook selectief. De meeste topics kunnen aan de hand van het analytisch kader in samenvattende vorm worden uitgeschreven, terwijl eventueel alleen nog voor specifieke topics en fragmenten waarin samenhangen naar voren komen volledige transcripties worden gemaakt.

De analyse is hier voornamelijk reflectie aan de hand van specifieke selecties uit het materiaal, overzichten en tabellen. Er zijn verschillende strategieën als hulpmiddel om de

kern op het spoor te komen. Naast memo-onderzoek en literatuurstudie (los van het interviewmateriaal) kan men de profielmemo's van de respondenten een voor een bestuderen, om na te gaan wat per respondent de kern lijkt te zijn die het een en ander met elkaar in verband brengt. Daarnaast kunnen uit alle interviews die fragmenten geselecteerd worden, die samenhang tussen verschillende onderwerpen thematiseren. Ook kan gebruik gemaakt worden van allerlei reeds geproduceerde overzichten met scores van respondenten bij verschillende variabelen, zodat naar samenhang gezocht kan worden.

De reductiefase is vooral gericht op de uitwerking van de kern van de theorie. Veelal gaat het hier om een kernbegrip, waarmee de overige centrale begrippen zijn verbonden. In veel kwalitatief onderzoek heeft het kernbegrip de vorm van een typologie. In de reductiefase is de analyse gericht op de uitwerking van dit kernbegrip en op het onderzoeken van de relaties van dit begrip met andere begrippen. De belangrijkste analysehandeling is dan ook *selectief coderen*, waarbij de bovengenoemde opties worden toegepast. Het invoeren van codes, het maken van filters, het selecteren van segmenten en het maken van prints van selecties uit het materiaal zijn ook hier administratieve handelingen die de computer kan overnemen.

Het doel van deze analytische activiteiten is om naar de kern van de nieuwe theorie te zoeken door de relaties tussen de belangrijkste variabelen na te gaan. Een handig instrument in deze fase is een optie om grafisch te werken, zodat je een schema van de relaties tussen de variabelen kunt uitwerken en dat in het werkbestand kunt opnemen. Als zo'n grafische omgeving niet in het programma is opgenomen, kunnen dit soort grafische memo's in een speciaal tekenprogramma worden gemaakt, maar uiteraard voldoen papier en potlood ook goed!

Het zoeken naar belangrijke variabelen en hun relaties betekent dat men ook allerlei soorten van matrices en tabellen wil kunnen maken. Als het gaat om matrices hoeft men niet meteen te denken aan de bekende datamatrix van de kwantitatieve analyse; er wordt gewerkt met kwalitatieve matrices waarin de cellen met

kwalitatieve beschrijvingen zijn gevuld. Miles en Huberman (1993) geven een groot aantal voorbeelden van de soort matrices die er zijn en hoe je er mee om kunt gaan.

Een andere belangrijke analysehandeling in deze fase bestaat uit de systematische en grondige bestudering van de memo's die in de loop van het onderzoek met betrekking tot de analyse zijn geschreven. De daarin opgenomen ideeën moeten gerecapituleerd en samengevat worden om het kernbegrip en de relaties met andere begrippen te onderbouwen.

De reductiefase is afgerond wanneer het kernbegrip is geformuleerd en de samenhang met de centrale onderwerpen van het onderzoek duidelijk is.

Bij deze globale beschrijving van de analyse in de reductiefase kunnen de volgende aantekeningen worden gemaakt.

#### **a. Verandering van analyseeniveau**

In de voorgaande fasen is de onderzoeker in de analyse voornamelijk bezig geweest op het niveau van het onderzoeksmateriaal, waarbij fragmenten werden gecodeerd en werden vergeleken. Alleen in (profiel)memo's heeft de onderzoeker op het niveau van de respondent gewerkt. In de reductiefase is de situatie juist andersom: nu vindt de analyse voornamelijk op het niveau van de respondenten plaats. De relevantie van het uitgewerkte analysekader kan getoetst worden door aan de hand daarvan de interviews samen te vatten en na te gaan of daarmee alle relevante informatie uit de interviews is opgenomen.

#### **b. Aandacht voor theoretische literatuur**

De onderzoeker heeft in de reductiefase behoefte aan een overzicht van theoretische invalshoeken met betrekking tot het onderzoeksveld om eventueel aan gangbare theorieën het kernbegrip te kunnen ontleenen. Daarnaast is de onderzoeker gericht op de formulering van het kernbegrip en ook hierbij kan de theoretische literatuur goede suggesties aandragen.

### **2.5 De integratiefase**

De integratiefase is erop gericht de theorie precies uit te werken, de relaties tussen het kernbegrip en de andere concepten in de theorie op

het materiaal te toetsen en een antwoord te formuleren op de probleemstelling en dit met onderzoeksmateriaal te illustreren. De analyse heeft dus het karakter van een systematisch onderzoek, waarin een vaststaand analytisch kader wordt gehanteerd.

De selectie van respondenten is gericht op een zodanige aanvulling op de reeds onderzochte gevallen, dat aan de doelstellingen van het onderzoek kan worden voldaan. Soms betekent dit dat naast de specifieke, extreme of bijzondere gevallen die in de eerste fasen op theoretische gronden zijn geselecteerd, juist nu aanvullend de aandacht is gericht op doorsnee gevallen, zodat het hele spectrum aan voorkomende gevallen in de onderzoeksgroep aanwezig is. Ook kan het onderzoeksdoel vereisen, dat aanvullend op aselechte wijze een flink aantal gevallen wordt geselecteerd om een representatief beeld van de populatie te kunnen schetsen.

De interviewer gaat in deze fase zeer gericht te werk. Weliswaar kunnen de gesprekstopics vereisen dat de respondent veel ruimte krijgt om het eigen perspectief naar voren te brengen, maar de interviewer is er met name op gericht bij elk gespreksonderwerp de respondent volgens het analytisch kader te positioneren. Het doorvragen is er dus op gericht duidelijk te krijgen waar de respondent thuis hoort.

Nog steeds kan de onderzoeker ervoor kiezen om ook deze gesprekken op tape op te nemen, maar de transcriptie van het materiaal is beperkt tot samenvattingen van wat gezegd wordt in termen van het analytisch kader.

De uiteindelijke analyse gebeurt op de profielmemo's van alle respondenten, die samenvattingen bevatten in termen van het analytisch kader. Hierbij kan voor elk geval worden getoetst of het kernbegrip in staat is de kern van het perspectief van de respondent weer te geven en hoe over de hele onderzoeksgroep de samenhang van het kernbegrip met de andere centrale begrippen eruitziet.

Deze fase behelst de uitwerking van de theorie in de zin dat de observatie-eenheden (respondenten) worden beschreven en gekarakteriseerd met behulp van de variabelen en relaties die tezamen de nieuwe theorie vormen. Matrices, tabellen, samenvattingen en allerlei soorten overzichten zijn de hulpmiddelen om pro-

fielmemo's en tabellen te produceren op het niveau van de nieuwe theorie. Een andere mogelijkheid is om via de computer quasi-kwantitatieve matrices te maken, waarin bijvoorbeeld staat aangegeven of een bepaalde code in het materiaal van een respondent of een document voorkomt, of in hoeveel segmenten van een document of interview een bepaalde code voorkomt. Deze matrices hebben een zodanig algemeen format, dat zij ingelezen kunnen worden in speciale programma's voor kwantitatieve analyse, zoals bijvoorbeeld een homogeniteitsanalyse (Homals) of correspondentieanalyse. Deze analyses zijn bedoeld om een beter beeld te krijgen van relaties tussen variabelen die zinvol gekwantificeerd kunnen worden, en zo bijvoorbeeld de eenheden (respondenten/documenten) in categorieën onder te brengen.

De laatste stap bestaat uit het uitschrijven van de theorie in een onderzoeksrapport. De basis hiervoor kan gevonden worden in de vele methodische en theoretische memo's die de onderzoeker in de loop van het onderzoek heeft geschreven.

Bij de analyse in deze fase zijn de volgende aantekeningen van belang.

#### **a. Herinterviews met 'oude' respondenten**

Omdat op dit moment in het onderzoek het analytisch kader van de onderzoeker pas definitief is geworden, kan nu ook duidelijk worden welke de beperkingen en lacunes zijn in de vraaggesprekken die in de eerste fase zijn gevoerd. Dit betekent dat deze respondenten aanvullend ondervraagd moeten worden. Dit kan belangrijke inzichten opleveren. Enerzijds zal blijken dat de respondenten over de gesprekstopics niet in precies dezelfde woorden zullen spreken, wat duidelijk kan maken of het analytisch kader deze 'alledaagse variatie' overstijgt. Anderzijds kan duidelijk worden dat sommige respondenten met betrekking tot de onderzochte onderwerpen van idee zijn veranderd. De vraag is dan of vanuit de theorie kan worden begrepen waarom die verandering heeft plaats gevonden.

#### **b. Analyseren is schrijven**

De meeste stappen in de analyse zijn op een of andere manier in schrijfproducten vastgelegd.

Daar kan de onderzoeker bij het uitschrijven van het onderzoeksrapport gebruik van maken, zoals de vele methodische en theoretische memo's, maar ook de profielmemo's van respondenten of gecodeerde fragmenten uit het onderzoeksmateriaal, die als illustratie kunnen dienen. Het schrijven van het eindrapport vraagt om nauwkeurige formuleringen, onderbouwingen van onderzoeksstappen en gefundeerde interpretaties van het onderzoeksmateriaal, die met elkaar consistent moeten zijn. In die zin is het schrijven van het eindrapport als laatste stap in de analyse te zien.

### 3 Slot

In het voorgaande hebben we de systematiek van de analyse in de gefundeerde theoriebenadering en de ondersteuning van de computer daarbij geschetst. De stappen in de analyse zijn daarbij alleen globaal aan de orde gekomen; de specifieke richtlijnen per stap zijn niet nader uitgewerkt. Ook zijn niet alle opties die het computerprogramma geeft, geïllustreerd (zoals het opstellen van matrices en het analyseren op woordniveau). Desondanks mag duidelijk zijn geworden dat de computer in alle fasen van de analyse belangrijke hulpfuncties kan vervullen.

Men zou een andere werkwijze dan de gefundeerde theoriebenadering kunnen toepassen, maar dat levert hoogstens andere paden op door het materiaal. Men kan een ander computerprogramma gebruiken, maar dat is zoiets als andere schoenen aantrekken. De basishandelingen met de computer zijn altijd min of meer hetzelfde. De computer kan de systematiek in de analyse bevorderen. In dit artikel hebben we laten zien, dat de analyse van kwalitatief materiaal gekenmerkt kan worden als een zoektocht langs analytisch kader en vraagstelling enerzijds en het empirisch materiaal anderzijds. De ondersteuning van de computer kan ertoe bijdragen, dat deze zoektocht meer systematisch en meer gestructureerd plaats vindt. Daarvoor is het wel noodzakelijk, dat het computerprogramma goed aansluit op de stappen die vanuit methodologisch perspectief noodzakelijk zijn.

Overigens moet men de mogelijkheden van de computer niet met bruikbaarheid voor de analyse verwarren. Hoewel de ontwikkelingen in de programmatuur steeds meer mogelijkhe-

den binnen bereik brengen zoals hypertext-achtige werkwijzen en de manipulatie van foto- en filmmateriaal, betekent dat tegelijkertijd dat de programma's zo algemeen worden (Ask Sam, Atlas-ti) dat de onderzoeker zelf zijn weg moet zoeken en eigen routines binnen het programma moet ontwikkelen. Wanneer ontwerpers zich bij het ontwikkelen van computerprogramma's voor de kwalitatieve analyse meer laten leiden door de technische mogelijkheden die de computer biedt dan door de context van een stapsgewijze analyseprocedure, kan het gebruik van de computer makkelijk tot valkuil verworden, waarbij toeters en bellen de systematiek in de analyse naar de achtergrond doen verdwijnen.

### Noten

- 1 Wij wijzen er op, dat het begrip analyse in het interpretatieve onderzoek veel breder wordt opgevat dan in andere vormen van onderzoek gangbaar is. Dit wordt in deze paragraaf toegelicht.
- 2 Coderen moet in deze context niet verward worden met het coderen dat plaatsvindt in kwantificerend onderzoek, waarbij vooraf bepaalde categorieën worden toegekend aan de antwoorden van de respondent. Binnen de context van interpreterend onderzoek wordt onder coderen verstaan de onderzoekshandeling waarbij de onderzoeker trefwoorden bedenkt die de inhoud van tekstfragmenten typeren; vervolgens worden deze trefwoorden of codes toegekend aan het betreffende fragment.
- 3 Binnen Kwalitan worden vier soorten memo's onderscheiden, elk met een eigen functie. In begrippenmemo's wordt de omschrijving van begrippen (codes, variabelen, dimensies) bijgehouden. Profielmemo's bevatten een beschrijving (profiel) van een respondent in termen van de begrippen die in de analyse ontwikkeld zijn. Theoriememo's bevatten theoretische overwegingen, zoals hypothesen en het analysekader-in-wording. Methodememo's, tot slot, bevatten beslissingen en verantwoordingen met betrekking tot het analyseverloop.

## Literatuur

- Fielding, N.G., & Lee, R.M. (1991). *Using computers in qualitative research*. London: Sage Publications.
- Glaser, B.G., & Strauss, A.L. (1967). *The Discovery of grounded theory*. Chicago: Aldine.
- Hijmans, E.J.S. (1994). *Je moet er het beste van maken*. Nijmegen: ITS.
- Hijmans, E.J.S. (1995). Kwalitatief survey. In: Hüttner, H., Renckstorf, K. & wester, F. *Onderzoekstypen in de communicatiewetenschap*. Houten: Bohn Stafleu Van Loghum.
- Kuyper, M. (1993). *Op de achtergrond. Een onderzoek naar de problemen van partners van patiënten met een chronische ziekte*. Amsterdam: Thesis publications.
- Maso, I. (1987). *Kwalitatief Onderzoek*. Meppel: Boom.
- Miles, M.B. and Huberman, A.M. (1984, 1993). *Qualitative Data-analysis. A sourcebook of new methods*. Sage: London.
- Muhr, Th. (1992). *Atlas / ti* (computer program). Berlin.
- Peters, V. (1995). *Kwalitan. A support program for qualitative analysis*. Version 4.0. (computer program). Nijmegen: University of Nijmegen.
- Peters, V. & Wester, F. (1995). An introduction to the principles and practice of our method of analysis. In: Pieterse, H.J.C. (ed.) *Desmond Tutu's Message. A qualitative Analysis*. Kampen (NL): Kok.
- Richards, T., & Richards, L. (1990). *Nudist 2.0: User's manual*. Melbourne: Replee.
- Spradley, J.P. (1979). *The Ethnographic Interview*. New York: Holt Rinehart Winston.
- Spradley, J.P. (1980). *Participant Observation*. New York: Holt Rinehart Winston.
- Strauss, A.L. (1987). *Qualitative Analysis for Social Scientists*. New York: Cambridge Univ. press.
- Strauss, A.L., & J. Corbin (1990). *Basics of Qualitative Research; Grounded Theory, Procedures and Tactics*. Sage; London.
- Seidel, J.V. (1988). *The Ethnograph version 3.0*. (computer program). Littleton: Qualis Research Associates.
- Weizman, E.A., & Miles, M.B. (1994). *Computer programs for qualitative analysis*. Thousand Oaks, CA: Sage.
- Wester, F. (1995). *Strategieën voor Kwalitatief Onderzoek*. Bussum: Coutinho (1e ed.1987).

## Auteurs

**Fred Wester** is universitair hoofddocent bij de vakgroep Methoden van de Faculteit Sociale Wetenschappen van de Katholieke Universiteit Nijmegen. Postbus 9104, HE Nijmegen.  
E-mail: f.wester@ped.kun.nl.

**Vincent Peters** is universitair docent bij de vakgroep Methoden van de faculteit Beleidswetenschappen van de Katholieke Universiteit Nijmegen. Postbus 9104, HE Nijmegen.  
E-mail: v.peters@ped.kun.nl

## Abstract

### **Qualitative analysis of interview data. Phases, steps and computer support in intensive research**

**F. Wester & V. Peters.** Pedagogische Studiën, 1999, 76, 117-131.

In this article a description is given of the design and the execution of the qualitative analysis of research material. At first, the characteristic cycle of observation – analysis – reflection, that is run through several times, is clarified. After that, the phases of the qualitative analysis are elaborated from the perspective of the grounded theory approach; the most important steps of analysis and the bottlenecks and pitfalls are pointed out. In addition, the way the computer can support this kind of analysis is illustrated by means of the program Kwalitan.

# Computerintensieve data-analyse: Bootstrap- en randomisatiemethoden in de statistiek\*

J.J. Hox

## Samenvatting

Dit artikel beschrijft twee computerintensieve methoden voor inductieve statistiek: bootstrappen en randomisatietoetsen. Bootstrappen houdt in dat uit de voorliggende gegevens een groot aantal steekproeven wordt getrokken, waarna in iedere bootstrapsteekproef de gebruikte statistische grootheid wordt berekend. De variabiliteit van de uitkomsten van de bootstrapsteekproeven wordt beschouwd als een indicatie van de steekproefvariatie in de echte steekproef. De bootstrapmethode wordt gebruikt om de standaardfout te schatten en daarmee een toets uit te voeren of betrouwbaarheidsintervallen te construeren. De randomisatietoets dient om verschillen tussen twee of meer groepen te toetsen. De gegevens van de twee groepen worden gecombineerd en vervolgens vele malen op een aselechte wijze in twee groepen verdeeld, waarna het verschil van de betreffende statistische grootheid in die twee groepen wordt berekend. De verdeling van de verschillen tussen twee toevallig samengestelde groepen wordt gebruikt als nulverdeling voor het daadwerkelijk geobserveerde verschil tussen de twee groepen. Bootstrap- en randomisatieprocedures zijn vooral bruikbaar in situaties waarin een klassieke statistische toets niet beschikbaar is, of wanneer ernstig aan de geldigheid van de aannamen van de klassieke toets getwijfeld wordt. Daarmee kunnen computerintensieve statistische methoden gezien worden als een alternatief voor de klassieke non-parametrische toetsen.

## Inleiding

Op het terrein van de statistische analyse heeft zich de laatste jaren, op een voor inhoudelijke onderzoekers vrijwel ongemerkte, bijna stiekeme manier, een nieuwe tak van statistiek ontwikkeld. Deze 'nieuwe statistiek' bestaat uit een verzameling van methodes, die ik hier samenvat met *computerintensieve statistiek*. Het kenmerkende van al deze methoden is dat er bij de bepaling van de significantie of het betrouwbaarheidsinterval niet gebruik wordt gemaakt van de klassieke statistische modellen voor de verdeling van de toevallige fouten, zoals een model met een normale verdeling. In plaats daarvan wordt het toevalsproces met de computer gesimuleerd. Deze simulatie wordt enige honderden tot duizenden malen herhaald, en de variaties in de uitkomsten van deze simulaties vormen een gesimuleerde empirische verdeling van de toevallige fouten.

Het statistisch jargon spreekt bij deze nieuwe toepassingen van resamplingmethoden, bootstrappen, randomisatie- en permutatietesten (zie o.a. Efron & Tibshirani, 1993; Mooney & Duval, 1993; Noreen, 1989). De wijze van werken laat zich het beste toelichten aan de hand van een voorbeeld. Bij wijze van illustratie volgt om te beginnen een voorbeeld van de zogenaamde *bootstrap*-methode.

## 1 Bootstrappen

Stel dat onderzoekers een vragenlijst willen ontwikkelen, bijvoorbeeld een korte attitude-schaal bestaande uit zeven vragen. Het is een goed gebruik om in een vooronderzoek voor zo'n schaal te bepalen, hoe betrouwbaar deze meet. Tabel A in de Bijlage bevat de (gefin-geerde) gegevens van een eerste vooronderzoek met vijftien personen. Bij deze proefafname blijkt de coëfficiënt alfa gelijk te zijn aan 0.74.

\* Met dank aan G.L.H. van den Wittenboer en AD. Nijdam voor hun commentaar bij een eerdere versie.



Doorgaans wordt een alfa van 0.70 of hoger beschouwd als een indicatie dat een schaal voldoende betrouwbaar meet. Een betrouwbaarheid van 0.74 lijkt dus voldoende. Het probleem is echter dat de proefgroep erg klein is, zodat we moeten aannemen dat er flinke steekproeffluctuaties in het spel kunnen zijn. Wat we nodig hebben is een betrouwbaarheidsinterval voor de coëfficiënt alfa. Hiervoor is een statistisch model ontwikkeld door Feldt (Feldt, 1965; Feldt, Woodruff & Salih, 1987; zie ook Van den Brink & Mellenbergh, 1998). De procedure van Feldt levert voor onze schaal een 95% betrouwbaarheidsinterval op van 0.47-0.90. Een nadeel van de procedure van Feldt is dat deze aanneemt dat de items normaal verdeeld zijn met gelijke variantie. Bovendien geldt het betreffende model pas bij grote aantallen vragen en bij grote steekproeven. Zeven vragen en vijftien respondenten lijken daarvoor beide aan de lage kant.

Een 95% betrouwbaarheidsinterval voor de coëfficiënt alfa kan echter ook worden vastgesteld met de *bootstrap*-methode. De bootstrapmethode is gebaseerd op de logica waarop statistische toetsen zijn gebaseerd. Die logica gaat ervan uit dat als we verscheidene malen een steekproef trekken uit een populatie, we vanzelfsprekend niet steeds precies dezelfde groep respondenten zullen krijgen, en daarom ook niet steeds dezelfde uitkomsten zullen vinden. De uitkomsten zullen van steekproef tot steekproef variëren en we gebruiken een mathematisch statistisch model om te schatten hoe groot de variatie is die we mogen verwachten. Op basis van aannamen voor de verdeling kunnen vervolgens betrouwbaarheidsintervallen bepaald worden en statistische toetsen worden uitgevoerd. De procedure van Feldt is gebaseerd op zo'n statistisch model.

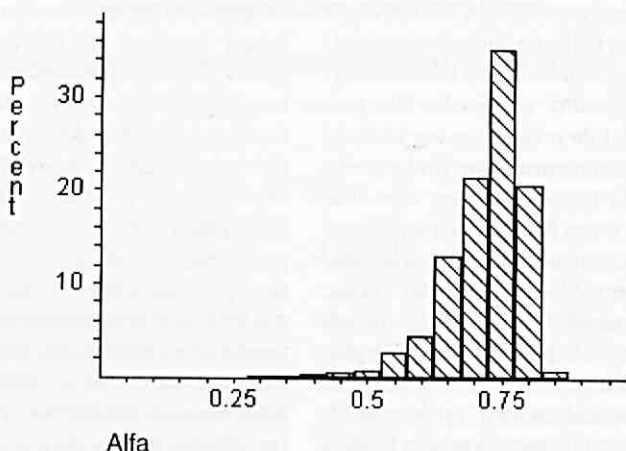
In plaats van een statistisch model te gebruiken, kunnen we natuurlijk ook duizendmaal met teruglegging een steekproef trekken, de gewenste uitkomsten berekenen, en bekijken hoe groot de variatie in uitkomsten daadwerkelijk is. In de praktijk is dat veel te bewerkelijk en bovendien te duur, maar in principe zou het kunnen. De bootstrapmethode bestaat uit een procedure die hier op lijkt. In plaats van duizendmaal een nieuwe steekproef te trekken uit de gehele populatie, trekt de computer duizendmaal een steekproef uit de reeds verzamel-

de gegevens. In het voorbeeld van de attitude-schaal wordt uit de voorliggende steekproef van vijftien respondenten duizendmaal met teruglegging een nieuwe steekproef van vijftien personen getrokken, en op die duizend nieuwe steekproeven wordt vervolgens duizendmaal de coëfficiënt alfa berekend. De variabiliteit van die duizend uitkomsten wordt gezien als een indicatie voor de grootte van de steekproeffluctuaties, zodat op basis hiervan een 95% betrouwbaarheidsinterval kan worden berekend en statistische toetsen kunnen worden uitgevoerd. Net als indertijd de Baron van Münchhausen trekken we ons aan onze laarzen het statistische moeras uit, vandaar de Engelse naam: *bootstrappen*.

Zoals hierboven is gesteld, trekken we de bootstrapsteekproeven met teruglegging. Telkens wanneer een respondent in de bootstrapsteekproef getrokken is, wordt deze weer teruggelegd bij de oorspronkelijke gegevens. De bootstrapsteekproeven verschillen dus van elkaar doordat sommige respondenten in een steekproef ontbreken, terwijl andere weer een aantal keren voorkomen. Omdat bij de bootstrap steeds opnieuw steekproeven uit de oorspronkelijke steekproef worden getrokken, heet deze strategie ook wel *resampling*.

Een resampling benadering van de betrouwbaarheid van coëfficiënt alfa houdt in dat we uit onze gegevens duizend bootstrapsteekproeven trekken. Vervolgens berekenen we duizendmaal de coëfficiënt alfa, en kijken bij de gevonden duizend alfa's welke de bovengrens en de ondergrens van de middelste 95% aanduiden. Het resultaat is dan een 95% bootstrapbetrouwbaarheidsinterval.

Figuur 1 geeft een histogram van de verdeling van coëfficiënt alfa in duizend bootstrapsteekproeven. Het 95% betrouwbaarheidsinterval volgens de bootstrapmethode loopt van 0.54 tot 0.82. Het percentage alfa's dat onder de kritische grens van 0.70 valt is 30%. De kans dat de betrouwbaarheidscoëfficiënt alfa in de populatie kleiner is dan 0.70, is aanzienlijk. Het betrouwbaarheidsinterval volgens de bootstrapmethode is kleiner dan het betrouwbaarheidsinterval volgens het model van Feldt (Feldt, 1965). In dit geval levert de klassieke statistische toets vermoedelijk een te grote standaardfout op. Aan het histogram van de



Figuur 1. Histogram van de waarden voor coëfficiënt alfa in 1000 bootstrapsteekproeven

bootstrap alfa's is goed te zien dat de steekproevenverdeling van alfa niet symmetrisch is. Vooral de scheve verdeling zorgt ervoor dat de kans op een fors lagere alfa bij vervolgonderzoek niet uitgesloten moet worden.

## 2 Randomisatie- en permutatietoetsen

Andere computerintensieve methoden zijn de zogenaamde randomisatie- of permutatietoetsen. Deze sluiten nauw aan bij de gedachten-gang in experimenteel onderzoek. Stel dat de gegevens van Tabel A verzameld zijn met behulp van een schriftelijke vragenlijst en we vragen ons af of dataverzameling door computerondersteund enquêteren meer betrouwbare gegevens zou geven (vgl. het artikel van Kef & Van Hattum in dit nummer). Volgens de beste principes van het experimentele onderzoek verzamelen we 50 respondenten, waarvan we er via loting 25 door de computer laten onder-vragen en de andere 25 een schriftelijke vragenlijst laten invullen. De computergestuurde interviewmethode levert een alfa op van 0.91, de vragenlijstmethode van 0.76. Mogen we nu concluderen dat de computergestuurde onder-vraging tot betrouwbaarder resultaten leidt?

Ook voor deze vraag is een statistische toets beschikbaar (Feldt, 1969; Hakstian & Whalen, 1976). Wanneer we deze toets toepassen, blijkt het verschil tussen de alfa's van beide groepen significant te zijn. De overschrijdingskans is

0.03, hetgeen wil zeggen dat de kans klein is dat dit soort verschillen op basis van toeval zijn ontstaan. We concluderen dat de wijze van vragen stellen invloed heeft op de betrouwbaarheid van de schaal.

De toets voor het verschil tussen twee alfa's berust op dezelfde aannamen als de hiervoor genoemde toets van Feldt (1965). Bij voorbaat weten we dan dat niet aan de aannamen voldaan is. We kunnen de toetsing echter ook op een andere wijze uitvoeren, namelijk via een *permutatie-* ofwel *randomisatietoets*. Een randomisatietoets is terug te voeren op de logica van het experimentele onderzoek. De basisvorm van het experimentele onderzoek is, dat een proefgroep volgens een toevalsproces wordt verdeeld in een experimentele groep en een controlegroep. Vervolgens wordt het experiment uitgevoerd en wordt nagegaan of de experimentele groep verschilt van de controlegroep. Doordat de groepsindeling volgens toeval, oftewel *at random* is geschied, mogen we aannemen dat de beide groepen aan het begin van het experiment gelijk zijn. Wanneer we achteraf verschillen constateren, kunnen we die toeschrijven aan de experimentele manipulatie.

De rol van de statistische toets bij experimenteel onderzoek is na te gaan of een eventueel gevonden verschil reëel is. Immers, wanneer we de proefpersonen at random over de experimentele en de controlegroep verdelen, zullen deze beide groepen doorgaans niet exact gelijk zijn. In ons voorbeeld is het erg onwaarschijnlijk dat de coëfficiënt alfa in beide groe-

pen in alle decimalen gelijk zal zijn. Kleine, toevallig optredende verschillen zijn te verwachten. Ook hier zouden we de toetsing via een statistisch model in principe kunnen omzeilen door het experiment een groot aantal malen te herhalen, met steeds andere toewijzingen van personen aan groepen. Om allerlei redenen is dat niet erg praktisch. Wat we wel kunnen doen is de personen waarvan we gegevens hebben, in de computer op alle mogelijke manieren in twee groepen te verdelen, van beide groepen telkens de coëfficiënt alfa bepalen, en daarvan ten slotte het verschil berekenen. De verdeling van al die verschillen tussen twee alfa's is een indicatie van de mate van variabiliteit in het verschil die we op basis van toeval mogen verwachten. Wanneer het daadwerkelijk geobserveerde verschil, vergeleken met die referentieverdeling, uitzonderlijk groot is, concluderen we dat het geobserveerde verschil niet op toeval berust, maar reëel is. Analooq aan het 5% significantieniveau van de gebruikelijke statistische toets, kunnen wij afspreken dat we een gevonden verschil significant noemen, als het zich bij de 5% grootste verschillen bevindt.

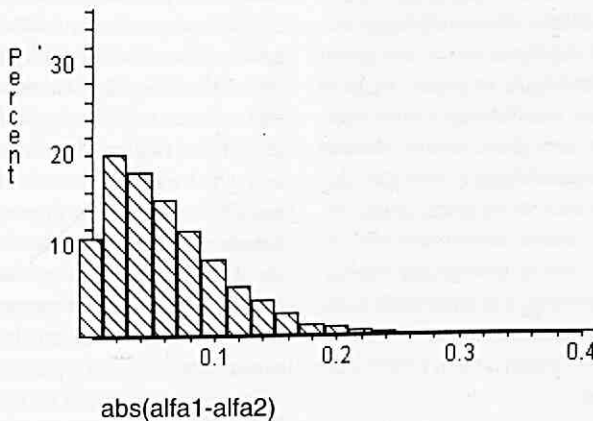
De zojuist beschreven toets heet een randomisatietoets, omdat deze het at random toewijzen van personen aan groepen nabootst. De randomisatie- of permutatietoets is al oud; hij is al in 1935 door Fisher beschreven. Nieuw is, dat de randomisatietoets tegenwoordig eenvoudig in de computer kan worden uitgevoerd en niet alleen meer voor eenvoudige grootheden zoals gemiddelden. Zolang het aantal personen

in de steekproef klein is, kunnen alle mogelijke manieren om twee groepen samen te stellen in de computer worden uitgerekend. Het aantal manieren om 50 personen over twee groepen te verdelen is echter onhanterbaar groot. Gebruikelijk is om dan bij de constructie van de referentieverdeling een steekproef te trekken en bijvoorbeeld duizend random toewijzingen uit te voeren en door te rekenen. Dit heet een approximatieve randomisatietoets.

Figuur 2 bevat het histogram van de verdeling van de absolute verschillen tussen twee alfa's bij duizend at random samengestelde groepen uit de steekproefgegevens in Tabel B van de bijlage. Het verschil tussen de betrouwbaarheidscoëfficiënten in de twee groepen bedraagt 0.15. Het percentage gevonden verschillen van 0.15 en groter is in de randomisatietoets 6%. Dat is iets groter dan de gebruikelijke kritische grens van 5%, en dus mogen we niet concluderen dat de beide groepen significant van elkaar verschillen. Omdat de overschrijdingskans in de randomisatietoets dicht bij de kritische significantiegrens van 0.05 ligt, is de randomisatietoets herhaald, nu met tienduizend aselekt samengestelde groepen. De overschrijdingskans in de randomisatietoets is nu 6.2%, en de conclusie blijft dat de nulhypothese hier niet kan worden verworpen.

### 3 Software

Computerintensieve methoden behoren nog niet tot het statistische standaardpakket. Recen-



Figuur 2. Histogram van de verschillen tussen de coëfficiënten alfa van twee groepen in duizend bootstrapsteekproeven.

Tabel 1  
Kruistabel van Tijdschrift en Soort probleemstelling

	PROBLEEMSTELLING (belangrijkste)					Totaal
	NAUWELIJKS AANWEZIG	EXPLO- RATIE	CONCRETE VRAAG	HYPO- THESE	INSTR. ONTWIKKELING	
TIJDSCHRIFTENCODI						
TS v. Opv., Vorming en Onderwijs	1	8	6	3	2	20
Pedagogisch Ts	4	9	11	9	4	37
Pedagogische Studiën		6	23	10	2	41
Kind en Adolescent	7	15	12	10	3	47
Ts v. Orthopedagogiek	11	15	13	11	11	61
Ts v. Onderwijsresearch	4	11	35	16	7	73
Ts v. Onderwijswetenschappen	3	6	21	2	3	35
Totaal	30	70	121	61	32	314

te versies van het statistisch pakket SPSS (vanaf 7.0) bevatten al enkele van de hierboven besproken methoden (Mehta & Patel, 1996). De module 'exact tests' voegt aan de kruistabel en de meeste nonparametrische toetsen de mogelijkheid toe om een 'exacte test' uit te voeren. In feite gaat het hierbij om de hierboven besproken randomisatietoets. SPSS berekent bij kleine steekproeven voor alle mogelijke steekproefindelingen de betreffende toetsingsgrootheid, en bepaalt vervolgens waar de geobserveerde toetsingsgrootheid zich bevindt in deze referentieverdeling. De proportiewaarden van de toetsingsgrootheid in de gesimuleerde randomisatieverdeling die groter is dan de gevonden waarde van de toetsingsgrootheid, levert de overschrijdingskans op voor de toets. Wanneer alle mogelijke steekproefindelingen kunnen worden berekend, is dit een exacte toets. In veel gevallen zal deze equivalent zijn aan een nonparametrische toets waarvoor de exacte overschrijdingskans bekend is. Bij grote steekproeven is het aantal mogelijke groepsindelingen te groot. In zo'n geval rekent SPSS de waarden uit van de toetsingsgrootheid voor een groot aantal (default 10000) random groepsindelingen. Het gaat dan om een approximatieve randomisatietoets, welke door SPSS wordt aangeduid als de 'Monte Carlo'-test<sup>2</sup>. Dit is niet langer equivalent aan de overeenkomstige nonparametrische toets, omdat deze bij grote steekproeven de overschrijdingskans bepaalt via een (doorgaans normale) benadering.

deling van het soort vraagstelling aangetroffen in artikelen in een aantal Nederlandse wetenschappelijke tijdschriften. Deze tabel is afkomstig uit een vergelijking van Nederlandse wetenschappelijke tijdschriften door Hox en Van Peet (1998). In het oorspronkelijke artikel wordt deze tabel geïnterpreteerd zonder dat er een statistische toets is uitgevoerd om na te gaan of de ogenschijnlijke verschillen tussen de tijdschriften significant zijn. De reden daarvoor is duidelijk wanneer we naar de aantallen in de cellen kijken. Wanneer we een gewone chi-kwadraat-toets opvragen voor de gegevens in Tabel 1, dan geeft SPSS de waarschuwing dat veel cellen een verwachte frequentie hebben kleiner dan vijf en dat toepassing van de chi-kwadraat-kansverdeling voor de toets niet verantwoord is. Het is echter mogelijk een exacte toets op te vragen. Dan berekent SPSS alle mogelijke manieren om de 314 waarnemingen, bij vastliggende randtotalen, over de 35 cellen van de kruistabel te verdelen. Dit komt overeen met de nulhypothese dat de beide variabelen die de kruistabel definiëren, niet met elkaar samenhangen. Voor elk van de mogelijke random indelingen berekent SPSS een chi-kwadraat-waarde. De gevonden chi-kwadraat-waarde wordt vergeleken met de chi-kwadraat-waarden uit de exacte kansverdeling, en de proportie chi-kwadraat-waarden gelijk aan of groter dan de gevonden chi-kwadraat-waarde levert wederom de exacte overschrijdingskans.

Tabel 1 bevat een kruistabel met een onderver-

Tabel 2 is een weergave van de SPSS-uitvoer voor de exacte toets op de gegevens van

Tabel 2

Monte Carlo significantietoets gebaseerd op tienduizend gesimuleerde kruistabellen

	Value	df	Asymp. Sig. (2-sided)	Sig.	Monte Carlo Sig. (2-sided)		Monte Carlo Sig. (1-sided)		
					9%		9%		
					Confidence Interval		Confidence Interval		
				Lower Bound	Upper Bound	Sig.	Lower Bound	Upper Bound	
Pearson Chi-Square	47.474 <sup>a</sup>	24	.003	.003 <sup>b</sup>	.001	.004			
Likelihood Ratio	51.267	24	.001	.002 <sup>b</sup>	.001	.003			
Fisher's Exact Test	46.897			.002 <sup>b</sup>	.001	.002			
Linear-by-Linear Association	.140 <sup>c</sup>	1	.708	.713 <sup>b</sup>	.701	.724	.356 <sup>b</sup>	.343	.368
N of Valid Cases	314								

a 12 cells (34.3%) have expected count less than 5. The minimum expected count is 1.91.

b Based on 10 000 sampled tables with starting seed 2 000 000.

c The standardized statistic is .374.

Tabel 1. Bij deze steekproefgrootte blijkt het niet mogelijk alle mogelijke verdelingen van 314 waarnemingen over 35 cellen te bepalen. SPSS voert daarom een approximatieve randomisatietoets uit, gebaseerd op 10000 random verdelingen. Dit levert bij deze gegevens voor de standaard Pearson chi-kwadraat vrijwel dezelfde overschrijdingskans op als de asymptotische:  $p$  is 0.003. Dat is op zich niet verwonderlijk; de regel die SPSS toepast bij het vermelden van de waarschuwing (niet meer dan 20% van de cellen mag een verwachte frequentie hebben lager dan 5) is nogal conservatief. Bij de Monte Carlo-resultaten krijgen we bovendien nog een 99% betrouwbaarheidsinterval van de berekende  $p$ -waarde, die immers gebaseerd is op tienduizend trekkingen van random-verdelingen. Als we het 99% betrouwbaarheidsinterval van de randomisatie  $p$ -waarde te breed vinden, kunnen we de toets eventueel overdoen en dan meer dan tienduizend trekkingen specificeren. In ons geval zijn de resultaten nu reeds duidelijk; de nulhypothese dat de verdeling van soorten probleemstellingen over de tijdschriften gelijk is, moet worden verworpen.

Bootstrapmethoden zijn (nog) niet beschikbaar in SPSS. Ze zijn al wel beschikbaar in een aantal programma's voor complexe modellen. Zo bevatten de programma's Lisrel, Eqs en Amos (alle bedoeld voor het schatten van structurele modellen) mogelijkheden om een bootstrap uit te voeren. Bij Lisrel en Eqs moet de gebruiker zelf nog het gemiddelde en de standaardafwijking van de bootstrapverdelingen berekenen,

bij Amos doet het programma alles automatisch. Ook andere complexe programma's, zoals het multiniveau-analyseprogramma MlwiN, bevatten tegenwoordig bootstrapprocedures.

Voor het toepassen van bootstrapprocedures op variantieanalyses, regressieanalyses, en andere gebruikelijke analysetechnieken moet een speciaal programma gebruikt worden. Het expertisecentrum ProGamma in Groningen heeft daarvoor het programma BOJA (Boomsma, 1991), dat bootstrap-schattingen kan berekenen voor een groot aantal toetsen.

Een speciale vermelding verdient het programma Resampling Stats (Simon, 1997). Dit is in feite een programmeertaal, waarin de gebruikers zelf hun statistische analyse moeten programmeren. Dat blijkt in de praktijk mee te vallen, omdat het programma veel commando's bevat die speciaal voor resamplingmethoden bedoeld zijn. Zo zet het commando SHUFFLE X alle getallen in de kolom X in random volgorde. Voor het bepalen van de randomisatieverdeling van de correlatie tussen X en Y, onder de nulhypothese dat er geen verband is tussen beide variabelen, is het dus voldoende om X duizendmaal te SHUFFLEn, en van de duizend berekende correlaties (commando: CORREL X Y in r) de verdeling te bepalen. Voor complexe variantieanalyses is Resampling Stats niet erg geschikt, maar voor randomisatietoetsen en bootstrapverdelingen van eenvoudige statistische grootheden zoals (verschillen tussen) gemiddelden, medianen, en correlaties is het vrij eenvoudig hanteerbaar.<sup>3</sup>

## 4 Voor- en nadelen van bootstrap- en randomisatiemethoden

Bootstrap- en randomisatiemethoden bestaan al enige tijd. Ze worden dikwijls toegepast wanneer langs de weg van de mathematische statistiek geen oplossing kan worden gevonden voor een specifiek toetsingsprobleem. Een voorbeeld daarvan is een betrouwbaarheidsinterval voor het verschil tussen twee medianen. Ze worden ook veel toegepast, als aan allerlei aannamen van de statistische toets niet is voldaan. De gedachte daarachter is in principe eenvoudig. Wanneer bijvoorbeeld de geobserveerde variabelen een zeer vreemde verdeling hebben, zodat aan de gebruikelijke aanname van een normale verdeling niet voldaan is, dan zal deze vreemde verdeling in elke bootstrap-steekproef opnieuw opduiken. Het probleem wordt daarmee vanzelf in de berekeningen verdisconteerd.<sup>4</sup> Omdat de bootstrap- en randomisatieprocedures alle informatie gebruiken die een steekproef bevat, hebben ze doorgaans ook een onderscheidingsvermogen (power) dat weinig onderdoet voor de gebruikelijke parametrische toetsen. Zie voor een bespreking van de power van randomisatieprocedures Onghe-  
na (1994.) Bootstrap- en randomisatieprocedures volgen doorgaans de logica van de parametrische toets, zonder een beroep te doen op de aanname van een normale verdeling. Nonparametrische toetsen volgen doorgaans een sterk verschillende logica, en hebben bij normale verdelingen minder onderscheidingsvermogen. Computerintensieve methoden zijn daarom een aantrekkelijk alternatief voor de gebruikelijke nonparametrische toetsen, die in veel gevallen minder onderscheidingsvermogen hebben dan hun parametrische tegenhangers, en bovendien soms ook sterke verdelingsaannamen doen. Zo veronderstelt bijvoorbeeld de bekende Mann-Whitney-toets voor het verschil in gemiddelde tussen twee groepen weliswaar niet een normale verdeling, maar wel dat beide groepen een gelijkvormige verdeling hebben.

Bootstrap- en randomisatietoetsen lossen uiteraard niet alle statistische problemen op. Allereerst wordt alle informatie uit de voorliggende steekproef betrokken. De validiteit van de statistische conclusies hangt daarom volle-

dig af van de representativiteit van die steekproef. Dat is bij de klassieke statistiek overigens ook het geval. Men name bij kleine steekproeven kunnen bootstrapprocedures in de problemen komen, omdat de variabiliteit in de populatie onderschat wordt. Daarnaast zijn er enkele statistische grootheden, zoals de grootste waarde in een steekproef, die zich slecht lenen voor bootstrappen of randomisatietoetsen (Mooney & Duval, 1993, p. 60).

Het moge duidelijk zijn waarom hier gesproken wordt van computerintensieve methoden. Het berekenen van één enkele coëfficiënt alfa is betrekkelijk weinig werk, zeker op een steekproef van slechts vijftig personen. Het berekenen van duizend alfa's valt voor een computer ook nog wel mee. Wanneer het daarentegen gaat om bootstrappen of randomisatietoetsen bij complexe multivariate modellen voor grote steekproeven, dan staat zelfs een moderne PC enige tijd te rekenen. Naarmate de computers sneller en goedkoper worden, zullen dit soort methoden echter aan populariteit winnen. De hier gegeven voorbeelden werden bijvoorbeeld berekend op een kleine Pentium laptop waarop alle berekeningen hoogstens enkele minuten duurden.

## Noten

- 1 Procedures gebaseerd op het simuleren van toevalsprocessen, waaronder de randomisatieprocedures vallen, worden dikwijls aangeduid met Monte Carlo procedures, naar het bekende casino.
- 2 Een demonstratieversie van Resampling Stats is te verkrijgen op het Internetadres <http://www.statistics.com>. Informatie over BOJA, inclusief een demonstratieversie, is verkrijgbaar bij Program op het Internetadres <http://www.gamma.rug.nl>.
- 3 De hier beschreven bootstrap-methode wordt wel aangemerkt als de naïeve of nonparametrische bootstrap. Wanneer deze methode gebruikt wordt voor het simuleren van de passing van complexe modellen, zijn de resultaten doorgaans onbetrouwbaar (Bollen & Stine, 1992). Dan biedt de parametrische bootstrap uitkomst: de achtereenvolgende steekproeven worden daarbij niet uit de data getrokken, maar met behulp van de geschatte modelparameters uit een theoretische kansverdeling gegenereerd

(vgl. Van der Heijden, 't Hart & Dessens, 1997). De parametrische bootstrap maakt daarmee wel weer gebruik van de aanname dat de gegevens een theoretische kansverdeling volgen.

Peet, A.A.J. van, Wittenboer, G.L.H. van den & Hox, J.J. (1995). *Beschrijvende technieken*. Groningen: Wolters-Noordhoff.  
Simon, J.L. (1997). *Resampling: the new statistics*. Arlington, VA: Resampling Stats.

## Literatuur

- Bollen, K.A., & Stine, R.A. (1992). Bootstrapping goodness-of-fit measures in the context of structural equation models. *Sociological Methods & Research*, 21, 205-229.
- Boomsma, A. (1991). *BOJA. A program for bootstrap and jackknife analysis*. Groningen: ProGamma.
- Brink, W.P. van den, & Mellenbergh, G.J. (1998). *Testleer en testconstructie*. Meppel: Boom.
- Efron, B., & Tibshirani, R.J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Feldt, L.S. (1965). The approximate sampling distribution of Kuder-Richardson coefficient twenty. *Psychometrika*, 30, 357-370.
- Feldt, L.S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika*, 34, 363-373.
- Feldt, L.S., Woodruff, D.J., & Salih, F.A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11, 93-103.
- Hakstian, A.R., & Whalen, T.E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika*, 41, 219-231.
- Heijden, P.G.M. van der, 't Hart, H., & Dessens, J. (1997). A parametric bootstrap procedure to perform statistical tests in a LCA of anti-social behavior. In: J. Rost & R. Langeheine eds. *Applications of latent trait and latent class models in the social sciences* (pp. 196-208). Münster/New York: Waxmann.
- Hox, J.J. & Peet, A.A.J. van (1998). Van oude methoden en dingen die voorbij gaan. *Pedagogische Studien*, jubileumnummer 1998.
- Mehta, C.R. & Patel, N.R. (1996). *SPSS Exact Tests 7.0 for Windows*. Chicago: SPSS Inc.
- Mooney, C.Z. & Duval, R.D. (1993). *Bootstrapping. A nonparametric approach to statistical inference*. Newbury Park, CA: Sage.
- Noreen, E.W. (1989). *Computer-intensive methods for testing hypotheses: An introduction*. New York: Wiley.
- Ongheena, P. (1994). *The power of randomization tests for single-case designs*. Academisch proefschrift, Katholieke Universiteit Leuven.

## Auteur

**J.J. Hox** is werkzaam aan de Universiteit van Utrecht als hoogleraar methodenleer. Dit artikel is voor een deel gebaseerd op zijn oratie, die op de homepage te verkrijgen is. Cap. groep Methodenleer en Statistiek, FSW, UU. Postbus 50140, 3508 TC Utrecht. E-mail [j.hox@fss.uu.nl](mailto:j.hox@fss.uu.nl).  
Homepage <http://www.fsw.ruu.nl/ms/jh>

## Abstract

### **Computer intensive data analysis: bootstrap and randomization methods in statistics**

**J.J. Hox.** Pedagogische Studiën, 1999, 76, 132-141.

This article describes two computer intensive methods for deciding on statistical generalizability: the bootstrap and randomization tests. In the bootstrap method, repeated samples are drawn from the available data, and subsequently the statistic of interest is calculated in each bootstrap sample. The variability of the results in the bootstrap samples is taken as indicative for the sampling variability of that specific statistic in the population at hand. Thus, the bootstrap sampling distribution is used to derive confidence intervals and significance tests. The randomization test is used to test group differences with respect to a specific statistic. The available data, ignoring the grouping structure, are randomly divided into groups following the group structure in the sample. For each random assignment, the difference of the specified statistic is calculated across the groups. The set of outcomes of the random assignments is then used as a null distribution for the observed difference. Bootstrap and randomization procedures are especially useful if there is no classical statistical procedure for the statistic of interest, or if the distribution of the data follows an unusual distribution. In the latter case, computer intensive methods are an alternative to classical nonparametric procedures.

## Bijlage

### De (gefingeerde) gegevens

Tabel A  
Scores van 15 personen op 7 vragen

Personen	1	2	3	4	5	6	7	Totaal
1	5	4	5	6	7	6	2	35
2	1	2	3	2	4	1	7	20
3	4	3	4	5	4	3	7	30
4	1	1	1	1	1	1	6	12
5	2	2	2	2	2	2	5	17
6	7	7	7	7	7	7	1	43
7	1	2	3	2	2	1	6	17
8	6	5	6	4	6	7	2	36
9	2	1	4	1	2	6	7	23
10	3	2	3	2	3	3	7	23
11	5	6	4	6	7	6	2	36
12	4	4	5	4	6	5	1	29
13	2	2	2	2	2	2	6	18
14	5	4	3	5	6	5	1	29
15	7	6	5	7	6	5	2	38

N.B. Deze tabel is afkomstig uit Van Peet, Van den Wittenboer & Hox, 1995, p. 210).



Tabel B

Scores van 50 personen op 7 vragen, CASI vs. PAPI.

Personen	1	2	3	4	5	6	7	Totaal	Groep
1	5	6	5	5	5	5	6	37	1
2	6	6	6	6	6	6	6	42	1
3	5	3	3	4	4	5	6	30	1
4	6	5	4	6	4	5	6	36	1
5	6	5	6	5	5	4	6	37	1
6	5	6	5	5	5	6	6	38	1
7	5	5	5	5	5	6	6	37	1
8	6	6	5	6	4	5	6	38	1
9	5	5	7	5	4	5	6	37	1
10	6	5	6	5	5	4	6	37	1
11	4	5	6	4	5	4	7	35	1
12	2	3	3	2	4	3	5	22	1
13	3	4	3	4	4	4	6	28	1
14	6	5	6	5	5	5	7	39	1
15	5	4	6	6	6	4	6	37	1
16	3	3	1	2	2	2	6	19	1
17	6	5	4	5	4	5	6	35	1
18	6	5	4	5	4	6	7	37	1
19	3	4	3	3	4	3	6	26	1
20	5	3	5	3	4	5	6	31	1
21	6	6	5	6	5	5	7	40	1
22	2	3	3	3	4	2	6	23	1
23	1	3	2	3	1	3	6	19	1
24	4	6	5	4	4	4	6	33	1
25	2	2	3	3	3	3	6	22	1
26	3	4	2	3	2	5	7	26	2
27	4	4	8	2	3	5	7	33	2
28	6	2	4	3	5	4	4	28	2
29	3	6	3	2	1	3	7	25	2
30	3	4	5	1	4	5	7	29	2
31	1	3	3	1	2	3	7	20	2
32	3	5	4	5	5	5	6	33	2
33	3	4	2	4	3	3	7	26	2
34	6	7	3	4	5	5	7	37	2
35	4	3	4	5	4	5	7	32	2
36	7	4	4	4	3	5	7	34	2
37	5	5	6	3	7	5	6	37	2
38	5	4	2	4	3	5	7	30	2
39	5	4	6	4	5	4	7	35	2
40	2	2	0	2	1	3	5	15	2
41	3	4	2	0	2	2	6	19	2
42	5	7	6	6	3	7	2	36	2
43	3	2	2	2	2	2	6	19	2
44	6	7	4	6	5	7	7	42	2
45	3	3	4	3	7	3	7	30	2
46	4	4	3	3	5	9	7	35	2
47	3	4	4	3	5	3	6	28	2
48	0	1	2	3	1	2	6	15	2
49	0	4	3	3	6	3	7	26	2
50	2	4	1	3	2	2	6	20	2