

1. Een schaap in de bus?

Een onderzoek naar waarden van de Nederlander

J.D. OPPENHUISEN en D. SIKKEL

SAMENVATTING

In dit artikel wordt een overzicht gegeven van een recente inventarisatie van de waarden van de Nederlander. Hierbij komt niet alleen de manier waarop de inventarisatie is uitgevoerd aan de orde, maar ook hoe deze inventarisatie is verwerkt tot een werkbaar model voor het doen van marktonderzoek. Daarnaast wordt beschreven hoe de inventarisatie toegankelijk is gemaakt via een cd-rom; deze kan worden gebruikt als hulpmiddel om ideeën te genereren voor het koppelen van waarden aan producten. Verder wordt aandacht besteed aan de vraag of de waardeoriëntatie die mensen hebben samenhangt met leeftijd of met generatie en wordt door middel van een blik in de toekomst het verschil tussen deze beide benaderingen duidelijk gemaakt.

1. INLEIDING

Waarden spelen in iedere samenleving een belangrijke rol in het dagelijkse leven. Ze stellen mensen in staat om keuzes te maken, ze dienen als motivatie van ideeën en vormen een leidraad voor gedrag (Rokeach, 1973). Waarden zijn cultuurafhankelijk (Hofstede, 1991) en worden al in de vroegste jeugd aangeleerd (Engel, Blackwell & Miniard, 1995; Hofstede, 1991; Rokeach, 1973), terwijl eenmaal aangeleerde waarden maar moeilijk kunnen worden afgeleerd en vervangen door andere. Waarden zijn min of meer stabiel, maar kunnen langzaam met het verstrijken van de tijd veranderen (Abramson & Inglehart, 1995; Becker & Nauta, 1983).

Ook in het marktonderzoek spelen waarden een belangrijke rol. Zowel bij de communicatie met de consument als bij het in kaart brengen van de markt en het positioneren van het product wordt volop gebruik gemaakt van waarden. Veelal wordt daarbij de waardelijst van Rokeach gebruikt. Ze worden vaak min of meer automatisch opgenomen in vragenlijsten waarmee voorkeuren van consumenten worden onderzocht. Ook in Nederland is dat het geval. Door waardelijsten te gebruiken die niet speciaal voor Nederland zijn ontwikkeld, worden elementen uit andere culturen in het Nederlandse consumentenonderzoek ingebracht. De vraag is nu: kan dat wel?

Waar cultuurvermenging toe kan leiden moge blijken uit het volgende, waar gebeurde, verhaal. Voor autochtone Nederlanders is het een doodnormale zaak om hun hond in trein en bus mee te nemen, maar als een allochtone Nederlander zijn schaap per bus wil vervoeren ontstaat de volgende situatie:

“Op 27 maart vieren de moslims Aïd-el-kebir (het offerfeest). Voor dit feest moet een schaap of geit worden geslacht en de Nederlandse overheid eist dat dit in slachthuizen gebeurt. De meeste schapen worden bij het slachthuis verhandeld, maar ook worden ze nog wel bij een boer gekocht, zodat het beestje dan naar het abattoir vervoerd moet worden.

*Dat deed Abdul in ieder geval vorig jaar. En omdat hij in Marokko zelf ooit slachter was, en hij zich met een schaap goed op zijn gemak voelde, besloot Abdul met de bus naar het slachthuis te gaan. Nederlanders nemen toch ook gewoon hun hond mee in de bus? Dus kocht hij een strippenkaart en stempelde twee strippen voor hemzelf en twee voor zijn schaap. De buschauffeur evenwel was not amused. D’r uit met dat schaap! Abdul moppert na in de klas. Als een hond in de bus mag, waarom een schaap dan niet? Zijn klasgenoten zijn het met hem eens. Liever een schaap in de bus dan een hond. Want honden zijn onreine, vieze dieren.” NRC 13 maart 1999: uit *Schaap in de bus door Elizabeth Termeer.**

Toch hebben autochtone Nederlanders niet altijd problemen met schapen, want ze vinden het wel toelaatbaar, ja zelfs prachtig, als er met Kerstmis een zogenaamde ‘levende kerststal’ in nota bene de kerk aanwezig is. Hierin bevinden zich dan niet alleen schapen, maar vaak zelfs een ezel en bij gebrek aan een os een koe. Maar een schaap in de bus, dat kan niet.

In de, in Nederland nog steeds gebruikte, waardelijst van Rokeach komt de waarde ‘verlossing’ voor. Vijftig jaar geleden wist vrijwel iedere Nederlander dat daarmee vrijmaking van de erfzonde werd bedoeld. In de huidige Nederlandse cultuur is het begrip ‘erfzonde’ zo goed als onbekend en is de waarde verlossing in een vragenlijst even vreemd als een schaap in de bus. In een onderzoek waarbij waarden worden gebruikt zal met dit soort culturele verschillen rekening moeten worden gehouden. Het is niet mogelijk om waardeonderzoek uit andere culturen zonder meer in Nederland toe te passen.

Binnen het werkprogramma van de Stichting Wetenschappelijk Onderzoek Commerciële Communicatie (SWOCC) spelen waarden en hun functie in de reclame een belangrijke rol. Omdat er vrijwel geen recent fundamenteel onderzoek is gedaan naar waarden in consumentengedrag, en zeker niet binnen de Nederlandse cultuur, is het hier beschreven project uitgevoerd. Twee vragen hebben daarbij centraal gestaan.

- Welke waardeomschrijvingen zijn er in Nederland? (Vooralsnog is het onderzoek beperkt tot de autochtone Nederlanders)
- Hoe zijn deze waardeomschrijvingen in dimensies te groeperen?

De opbouw van dit artikel is als volgt: eerst wordt kort ingegaan op de theorieën over waarden en waardeonderzoek, vervolgens volgt een overzicht van de gebruikte inventarisatie methode en de resultaten. Daarna komt de toepassing van het gevonden model aan de orde en wordt een blik in de toekomst geworpen waarna nog enkele conclusies volgen.

2. ACHTERGRONDEN VAN WAARDEN EN WAARDE-ONDERZOEK

Hoewel er veel waardeonderzoek wordt uitgevoerd is het niet zo eenvoudig om aan te geven wat een waarde nu eigenlijk is. Er zijn zeer veel definities van waarden. Lautman (1971) heeft er in 1971 al meer dan 170 gevonden in 400 verschillende bronnen.

Schwartz & Bilsky (1987) hebben geconcludeerd dat de meeste definities een aantal overeenkomsten hebben. Volgens hen zijn waarden:

- a *concepts or beliefs*
- b *about desirable end states or behaviors*
- c *that transcend specific situations*
- d *guide selection or evaluation of behavior and events, and*
- e *are ordered by relative importance* (Schwartz & Bilsky, 1987, p. 551).

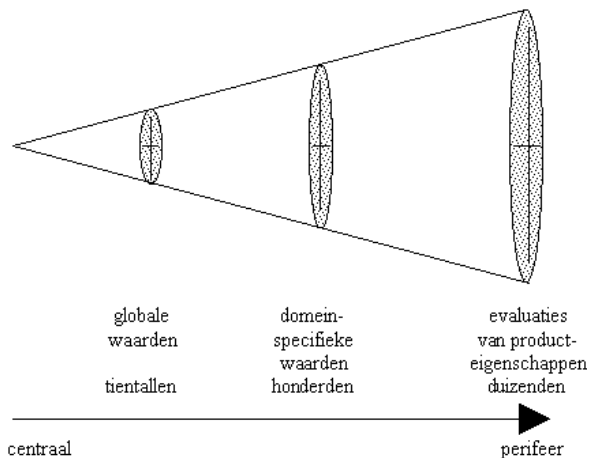
Een veelgebruikte waardelijst is die van Rokeach. De definitie die Rokeach heeft gebruikt is met bovenstaande elementen in overeenstemming

A value is an enduring belief that a specific mode of conduct or end-state of existence is personally or socially preferable to an opposite or converse mode of conduct or end-state of existence (Rokeach, 1973, p. 5).

Hij maakt onderscheid in 18 instrumentele- en 18 eindwaarden, waarbij de instrumentele waarden worden gebruikt om de eindwaarden te bereiken. Vinson, Scott en Lamont (1977) kennen een ruimer waardebegrip. Hun definitie, die ook past in de criteria van Schwartz en Bilsky, luidt:

“Values are centrally held cognitive elements which stimulate motivation for behavioral response. They exist in an interconnected hierarchical structure in which global values are related and connected to generalized consumption-related values which are, in turn, similarly associated with product attributes” (Vinson et al., 1977, p. 49).

Zij hebben een model ontworpen dat de relaties weergeeft tussen de universele waarden die mensen hebben (enkele tientallen), de domeinspecifieke waarden (enkele honderden) en de waarden die producteigenschappen beschrijven (vele duizenden).



Figuur 1. Model van samenhang tussen waarden. Vrij naar Vinson, Scott en Lamont (Sikkel, 1997).

De manier waarop een product door een persoon wordt beschreven hangt af van zijn globale en domeinspecifieke waarden. Zo zal de ene persoon kleding beschrijven als ‘zit lekker’ en ‘gaat lang mee’ terwijl voor de ander kleding ‘modieus is’ en ‘mooi staat’. De eerste persoon hecht veel waarde aan *duurzaamheid* en *comfort*, terwijl de tweede het kennelijk belangrijk vindt dat hij laat zien dat hij *bij de tijd is* en *er goed uit ziet*.

In dit onderzoek wordt de volgende definitie van een waarde gebruikt:

Een waarde is een toestand of gedragwijze die mensen nastreven (Oppenhuisen, 2000 p.87).

Deze definitie is afgeleid van die van Rokeach (1973) maar is ingekort, omdat er bij de inventarisatie van de waarden van de Nederlander duizenden uitspraken van respondenten op waarden moeten worden onderzocht. Hiervoor is het noodzakelijk om een korte, goed werkbaar definitie te gebruiken. Door het verkorten voldoet de definitie alleen aan de elementen a en b die door Schwarz en Bilsky (1987) zijn genoemd. Aan de voorwaarde dat een waarde specifieke situaties moet overstijgen (c) wordt voldaan door de aard van het onderzoek en komt in paragraaf drie nader aan de orde. De punten d en e kunnen als eigenschappen van een waarde worden beschouwd en kunnen in vervolgonderzoek gecheckt worden. De gebruikte definitie staat een ruimere inventarisatie toe dan de definitie van Rokeach zou toelaten, maar sluit daardoor meer aan bij de indeling die Vinson et al. (1977) gemaakt hebben van waarden.

3. METHODE VAN ONDERZOEK

Om waarden van mensen te achterhalen staan een aantal methoden open. De eerste en meest directe is het gewoon vragen aan mensen. Een probleem is daarbij: ‘hoe maak ik de respondent duidelijk wat een waarde is’. Het begrip ‘waarde’ is te abstract om er zonder meer aan een persoon naar te kunnen vragen. Daarnaast is het nog maar de vraag of mensen bereid zijn om hun waarden aan iemand te vertellen als ze zelf al weten welke waarden ze hebben (Becker & Nauta, 1983; Halman, 1991; Hofstede, 1991; Kraemer, 1980; Rokeach, 1973). Een andere methode is personen langdurig te observeren en dan uit het gedrag de waarden afleiden (Hofstede, 1991). Het bezwaar van deze methode is dat zij zeer tijdrovend is en in het algemeen toch als een inbreuk op de privacy wordt ervaren. Een derde manier is waarden vaststellen aan de hand van boeken die mensen lezen. Vooral kinderboeken zijn een rijke waardebron, omdat waarden in de vroegste jeugd worden aangeleerd. Posthouwer (1998) heeft op deze wijze waardeonderzoek uitgevoerd. Nog andere methoden kunnen zijn het analyseren van soaps (Van Driel, 2000), nieuwsuitzendingen, officiële documenten (Becker & Nauta, 1983), persoonlijke brieven, enzovoort. Echter boeken, soaps en de andere genoemde zaken worden door een selecte groep mensen gemaakt. De waarden die in deze stukken voorkomen zijn daardoor niet representatief voor de gemiddelde Nederlander.

3.1. Gebruikte methode

Voor dit onderzoek is gekozen om met behulp van diepte-interviews respondenten te ondervragen over allerlei onderwerpen die in het leven van bijna alle mensen voorko-

men. Hierbij komen onderwerpen aan de orde als: kinderen krijgen, naar school gaan, werken, vakantie, politiek en dood gaan. Voor het interviewen van respondenten is gekozen, omdat deze methode het dichtste komt bij mensen vragen welke waarden zij hebben (Halman, 1991). Om respondenten te stimuleren over bepaalde onderwerpen te praten is een fotoserie samengesteld die bestaat uit 22 foto's/ fotocollages die elk een bepaalde fase uit het leven van de mens bestrijken. Bij elke foto is een introductie vraag ontwikkeld, omdat het voor sommige respondenten moeilijk blijkt te zijn om spontaan over de foto's te vertellen, vragen als: 'Waarom willen mensen kinderen', 'Waarom willen mensen trouwen' (zie voor een volledig overzicht Bijlage 1). Daarnaast zijn er bij de foto's nog een aantal vragen gemaakt om bepaalde onderwerpen die weliswaar met de foto samenhangen, maar niet spontaan aan de orde komen te kunnen behandelen. Dit zijn vragen als: 'Waarom willen mensen een groot gezin', 'Waarom gaan mensen soms ook scheiden' (ook deze vragen staan in Bijlage 1). Deze standaardvragen hebben een aantal voordelen:

- ze maken het voor de respondent makkelijker om over een onderwerp te praten
- ze leveren standaardisatie van de interviews op
- ze geven de zekerheid dat bepaalde onderwerpen aan de orde komen
- ze maken het mogelijk het onderzoek te herhalen op dezelfde manier
- ze zorgen ervoor dat er minder foto's nodig zijn, omdat meerdere onderwerpen bij dezelfde foto aan de orde gesteld kunnen worden.

Ter controle wordt een tweede fotoserie ontwikkeld. De vragen blijven hetzelfde. De interviews zijn gebaseerd op de laddering methode (Reynolds & Gutman, 1988). Bij laddering wordt, nadat de respondent een antwoord heeft gegeven, steeds een 'waarom vraag' gesteld. Hiermee wordt steeds dieper op het onderwerp ingegaan, net zolang tot de respondent niet verder meer kan.

Voor het uiteindelijke onderzoek is gebruik gemaakt van 20 kwalitatieve interviews. De geïnterviewde respondenten waren gespreid naar regio, geslacht en opleiding. Voor leeftijd is gekozen voor jongeren (20-30 jaar) en ouderen (50-60 jaar). Dit is gedaan omdat uit het proefonderzoek is gebleken dat het waardepatroon geen verschil naar leeftijd te zien geeft. Om dat te kunnen testen is voor deze indeling gekozen.

De uitspraken van de respondenten worden in zijn geheel letterlijk uitgetypt. Met behulp van de definitie van een waarde, *een waarde is een toestand of gedragwijze die mensen nastreven*, worden uit de uitspraken waarden geselecteerd. De waarden bestaan zo veel mogelijk uit de letterlijke tekst van de respondenten. Hiervoor is om twee redenen gekozen:

- het is het meest objectief
- het geeft inzicht in de manier waarop respondenten met een verschillende achtergrond zich uitdrukken.

Dit laatste biedt aan de marketingwereld de mogelijkheid om het taalgebruik in campagnes en reclames aan te passen aan de doelgroep. De volledige interviews, de foto's, de gegenereerde waarden en de later uitgevoerde analyses worden opgenomen in een database welke is verschenen op cd-rom, het Waardenwoordenboek. Het gebruik van deze database heeft twee voordelen:

- Hiermee is het onderzoek volledig herhaalbaar en worden subjectieve beslissingen controleerbaar.
- De onderlinge relaties tussen uitspraken van respondenten, gegenereerde waarden,

getoonde foto's, gestelde hoofdvragen en kenmerken van respondenten kunnen op eenvoudige wijze zichtbaar worden gemaakt.

De database bevat 1372 omschrijvingen van waarden waarvan de verschillende contexten eenvoudig kunnen worden nagezocht. Voor het uitvoeren van consumentenonderzoek is het aantal van 1372 omschrijvingen uiteraard veel te hoog. Daarom worden deze omschrijvingen in groepen samengevoegd die met één waarde worden omschreven. Daarbij zijn de volgende criteria gebruikt:

- Synoniemen. Waarden die synoniem zijn worden samengevoegd onder een noemer.
 - Ambigüiteit. Waarden die voor meerdere uitleg vatbaar zijn worden ondergebracht bij soortgelijke waarden die dat niet of veel minder zijn.
 - Aantal malen dat een waarde voorkomt. Waarden die veel worden genoemd gaan in het algemeen boven waarden die een enkele keer worden genoemd, tenzij de veelgenoemde waarden ambigu zijn.
 - De omschrijving die in het Van Dale woordenboek staat.
- Uiteindelijk blijven er 160 waarden over die de andere waarden beschrijven.

Een voorbeeld is de waarde 'actief zijn'. Aan deze waarde zijn 18 omschrijvingen gekoppeld die door 'actief zijn' worden gerepresenteerd.

Waarden verbonden met **actief zijn**

bezig zijn met	dynamisch zijn	in het leven staan
jezelf vermaken	niet lui zijn	de wijde wereld ingaan
je niet vervelen	ondernemend zijn	plannen hebben
snel zijn	vlot zijn	iets uitvoeren
iets te doen willen hebben	iets omhanden hebben	energie hebben
het leven beleven	een bezigheid hebben	tijdverdrijf hebben

4. RESULTATEN

Ook de gereduceerde set van 160 waarden is nog te groot om te kunnen gebruiken in marketing of ander onderzoek. Daarom wordt het aantal waarden nog verder teruggebracht door ze met behulp van Principale Componenten Analyse (PCA) in kleinere eenheden in te delen. Om deze techniek te kunnen toepassen moeten eerst de afstanden tussen de waarden worden bepaald. Dit gebeurt door een representatieve steekproef uit de Nederlandse bevolking, i.c. het CentERpanel. Er zijn echter $160 * 159/2 = 12720$ afstanden te bepalen. Uit vooronderzoek is gebleken dat personen maximaal 70 afstanden kunnen beoordelen, daarna slaat de vermoeidheid toe en zijn de resultaten niet meer betrouwbaar. Daarom worden de afstanden zodanig over de panelleden verdeeld, dat elke afstand gemiddeld 12 maal voorkomt. Ieder panellid beoordeelt dan 64 afstanden. Door de waardecombinaties en de volgorde waarin de waarden voorkomen te variëren, kunnen de context- en volgorde effecten worden geminimaliseerd. Panellid 1 vergelijkt bijvoorbeeld de waarden 1 tot en met 8 met de waarden 9 tot en met 16, terwijl panellid 2 de waarden 2 tot en met 9 vergelijkt met de waarden 10 tot en met 17. Hierbij moet dan door middel van een cijfer tussen de 1 en 10 worden aangegeven of de verschillende waarden veel met elkaar overeenkomen (10) of juist helemaal niet (1).

Dus liggen de waarden ‘moeder zijn’ en ‘baas zijn’ dicht bij of juist ver uit elkaar. Op deze manier wordt uiteindelijk een matrix geconstrueerd van alle afstanden tussen de 160 waarden. Deze matrix vormt de basis voor de PCA.

4.1. Dimensies

Met PCA worden de dimensies in beeld gebracht die de belangrijkste tegenstellingen in de waarden weergeven. De eerste zes dimensies zijn goed interpreteerbaar in algemene termen. Binnen de hogere factoren zitten nog wel interpreteerbare waardeclusters, maar de benoeming van assen als geheel is niet meer mogelijk. In Tabel 1 staan de zes dimensies met steeds de tien waarden die de meest extreme factorscores op die dimensie hebben.

De factoren hebben gemeenschappelijk dat er steeds een element van vrijheid tegenover binding in zit: het object van de binding is echter per factor anders. Zo is het bij factor 1 de sociale relaties (al kan natuurlijk ook hang naar een carrière als binding worden opgevat), bij factor 2 is het de medemens, bij factor 3 traditie, bij factor 4 zekerheid, bij factor 5 gezinsleven en bij factor 6 andermans goedkeuring.

Verskillende dimensies uit het SWOCC-waardeonderzoek zijn goed vergelijkbaar met dimensies die in ander onderzoek gevonden worden. In Tabel 2 staat een overzicht van dimensies uit ander onderzoek, waarbij de overeenkomstige waarden uit het SWOCC onderzoek eveneens zijn aangegeven. Echter, in ander onderzoek worden nooit meer dan hooguit drie dimensies gevonden, met uitzondering van Hofstede, terwijl wij zes dimensies aantreffen, die, zo blijkt uit latere toepassing, allemaal een duidelijke en verschillende betekenis hebben (Oppenhuisen, 2000). (VNU heeft de dimensies gebruikt voor het in kaart brengen van een groot aantal tijdschriften en merken. Uit dat onderzoek is duidelijk gebleken dat het kunnen beschikken over meerdere dimensies voordelen heeft). Voor het marktonderzoek betekent dit dat een groter aantal dimensies voorhanden is om te differentiëren naar producten en merken. Niet alle dimensies zullen voor iedere markt relevant zijn, maar de kans dat er relevante dimensies worden gevonden, is aanzienlijk groter dan wanneer slechts met twee of drie dimensies wordt gewerkt.

4.2. Waardenwoordenboek (cd-rom)

Een belangrijk resultaat van het onderzoek is de waarden cd-rom, het Waardenwoordenboek. Zowel de foto's, de standaardvragen, de achtergrondvariabelen en de uitspraken van de respondenten, de 160 waarden die de overige waarden representeren, de resultaten van de analyses en een beschrijving van het onderzoek zijn opgenomen op deze cd-rom. Ook alle subjectieve beslissingen die gaandeweg het onderzoek genomen moesten worden (foto keuze, vragen, selectie van de waarden, de keuze van de 160 waarden) zijn vastgelegd. Het totale onderzoek is hiermee voor iedereen controleer- en herhaalbaar geworden. Daarnaast biedt de cd-rom de mogelijkheid om voor marketingdoeleinden die waarden te kiezen die het beste bij het product passen. De 160 waarden die gekozen zijn om de andere waarden te vertegenwoordigen staan in de thesaurus aangegeven met een asterisk (*) en worden daarom in het vervolg aangeduid met waarden*. In de thesaurus zijn de waarden* gekoppeld aan de waarden die zij vertegenwoordigen. Zo is de waarde ‘aan beeld voldoen’ gekoppeld aan de waarden* ‘bij de groep horen’, ‘erkenning krijgen’ en ‘gehoorzaam zijn’. Maar ook is te zien bij welke uitspraak, respondent en foto de waarde hoort.

Tabel 1. De interpretatie van de dimensies 1 tot en met 6 met de meest kenmerkende waarden.

Factor	1	2	3
<i>binding</i>	Relaties	Sociaal georiënteerd	Waarden van oudsher
↑	knuffelen vriendschap liefde gezelligheid sfeer vertederen iets voor iemand over- hebben een partner hebben een fijn gezin hebben gemoedelijk zijn	luisteren naar iemand begrip hebben voor behulpzaam zijn verantwoordelijk zijn mensenkennis hebben rekening houden met iemand serieus zijn samen werken wijsheid (iets oplossen)	vaderlands liefde deftig zijn trots stoer zijn aantrekkelijk zijn onsterfelijk zijn een held zijn roots (wortels) hebben aanzien hebben spiegelen aan
↓	eigenwijs zijn vasthoudend zijn je eigen weg gaan vakmanschap presteren ambities hebben doorzettingsvermogen macht fanatiek zijn hoger op komen	er verzorgd uitzien seksuele gemeenschap hebben een kick krijgen plezier hebben ontspanning hebben gezond zijn jong blijven/ zijn onbezorgd zijn aantrekkelijk zijn genieten	vakmanschap plezier hebben (iets) oplossen creatief zijn actief zijn ontspanning hebben iets maken tijd hebben leren een hobby hebben
<i>vrijheid</i>	Prestatie	Individueel georiënteerd	Waarden van nu
Factor	4	5	6
<i>binding</i>	Zekerheid	Gezinsleven	Conformistisch zijn
↑	netheid rijk zijn schoon zijn veiligheid rust luxe eigen plek hebben spaarzaam zijn in je onderhoud voorzien zekerheid	moeder zijn kinderen hebben knuffelen zorgen voor verwennen een partner hebben vertederen opvoeden seksuele gemeenschap hebben liefde	er verzorgd uitzien aantrekkelijk zijn aanzien hebben netheid deftig zijn gezag hebben beleefdheid schoon zijn baas zijn middelpunt zijn
↓	spanning (zelf gezocht) stoer zijn ruimdenkend zijn humor een held zijn eigen mening hebben boeiend zijn grenzen verleggen spontaan zijn uitdagend zijn	niets te kort komen onafhankelijk zijn eenvoud niet jaloers zijn humor tevreden zijn onbezorgd zijn rust gewoon zijn vrijheid	spaarzaam zijn alleen zijn (vrijwillig) eigen plek hebben vrijheid idealistisch zijn beschermen vaderlandsliefde opstandig zijn geloven je eigen weg gaan
<i>vrijheid</i>	Uitdaging	Vrijheid	Je eigen weg gaan

Tabel 2. Een vergelijking tussen de belangrijkste dimensies uit verschillende waardesystemen en de SWOCC- waarde-inventarisatie (vrij naar Hansen, 1998).

Onderzoek methode	Dimensie 1 ander onderzoek	Dimensie SWOCC	Dimensie 2 ander onderzoek	Dimensie SWOCC	Dimensie 3 ander onderzoek	Dimensie SWOCC
VALS	Modern ↔ traditioneel	3	Actie gericht ↔ principe gericht	4		
RISC	Openstaan ↔ weerstand tegen verandering	3	Ethisch ↔ hedonisme			
CCA	Verandering ↔ stabiliteit	3	Materieel ↔ spiritueel werken			
Grunert/Schwartz	Open voor verandering ↔ conserverend	3	Boven jezelf uitstijgen ↔ jezelf verbeteren	2		
Value-scope	Verandering naar modern ↔ stabiliteit	3	Pragmatisch ↔ loyaliteit		Materialisme ↔ humaniteit	1
Kompas	Modern ↔ traditioneel	3	Individueel georiënteerd ↔ sociaal georiënteerd	2		
Danish attitudes	Modern ↔ traditioneel	3	Individueel ↔ collectief	2		
Minerva	Modern ↔ traditioneel	3	Sociaal ↔ individueel	2		
PDA	Vitaliteit ↔ continuïteit	3	Prestatie ↔ harmonie	1		
Vyncke	Ik-gericht ↔ ander-gericht	2	Materialisme dimensie			
Quality Planning	Ik ↔ ander	2	Op buitenwereld gericht ↔ op innerlijk gericht		Norm aanpassend ↔ norm onafhankelijk gedrag	6
Hofstede*	Collectivisme ↔ individualisme	↔	Masculien ↔ feminien	1	Onzekerheidsvermijding	4

* De andere dimensies van Hofstede worden niet in het SWOCC-onderzoek teruggevonden en zijn om technische redenen niet in de tabel opgenomen.

4.3. Betrouwbaarheid en validiteit van het onderzoek

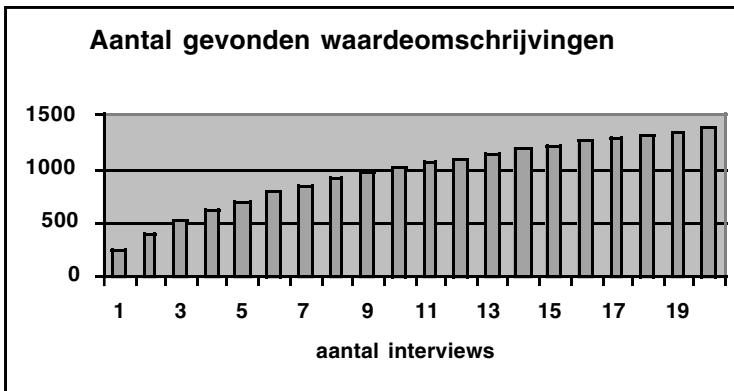
Gezien het kleine aantal interviews (20) en het geringe aantal keren dat een afstand is beoordeeld (12) is de vraag gerechtvaardigd hoe betrouwbaar deze inventarisatie is? Daarom wordt aan de validiteit en betrouwbaarheid hierna ruime aandacht besteed. Allereerst wordt gekeken naar de gebruikte stimuli, daarna staat het aantal interviews ter discussie en vervolgens wordt de stabiliteit van de gevonden dimensies bekeken.

4.3.1. Stimuli

Zijn de waarden die gevonden worden de waarden van mensen in het algemeen of de respondent in bijzonder of zijn het waarden die zijn gekoppeld aan toevallige details van de gebruikte foto's? Met andere woorden: in hoeverre hangen de uitspraken van de respondenten af van de gebruikte stimuli? Om dat te toetsen zijn twee verschillende fotoseries gebruikt. Om de series te kunnen vergelijken worden de waarden niet per stuk vergeleken, maar wordt een vergelijking gemaakt van de waardeconfiguratie als geheel. Als basis worden daarbij de zes factoren gebruikt. Per factor en per serie worden het gewogen gemiddelde en de gewogen standaardafwijking berekend (de wegingsfactor is het aantal keren dat een waarde in het onderzoek is genoemd). Wanneer de foto's invloed op de resultaten zouden hebben zou zich dat in verschillen in gemiddelde en standaardafwijking op de factoren moeten terugvertalen. Dit bleek niet het geval te zijn.

4.3.2. Volledigheid inventarisatie

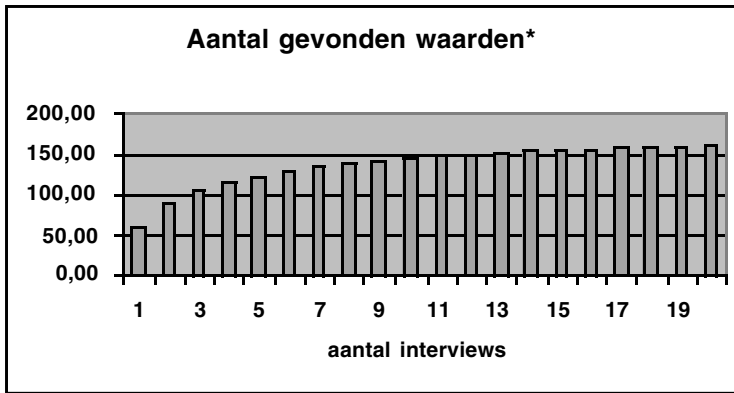
In de regel kan gesteld worden dat bij kwalitatieve interviews het aantal voldoende is als er geen nieuwe informatie meer wordt gevonden in nieuwe interviews. Deze regel kunnen we toepassen op het totaal aantal gevonden waarden en op de 160 waarden*, maar ook op de factoroplossingen.



Figuur 2. Het gemiddeld aantal gevonden waardeomschrijvingen als een functie van het aantal interviews.

In figuur 2 staat het gemiddeld aantal waarden die gevonden zijn bij een bepaald aantal interviews. Het gemiddelde is genomen over alle mogelijke interview combinaties (voor 5 interviews zijn dat $20!/(5! \times 15!)$ combinaties). Uit de figuur is af te lezen dat er in interview 19 en 20 nog steeds nieuwe waarden gevonden worden. Men kan dus aannemen dat er nog veel meer waarden te vinden zijn en dat de lijst van 1372 waarden nog verre van compleet is.

In figuur 3 is te zien dat het aantal waarden* nauwelijks meer toeneemt na interview elf. Op dat moment is het gemiddelde aantal waarden* 148. Na interview 17 stopt de groei helemaal. Deze op zichzelf positieve uitkomst zou een artefact kunnen zijn van



Figuur 3. Het gemiddeld aantal waarden* als een functie van het aantal interviews.

de subjectieve keuze van de onderzoekers om waarden* te selecteren die ‘gewone’ veel voorkomende woorden zijn. Daarom zijn er andere argumenten nodig om aan te tonen dat het aantal van 160 waarden* voldoende is om de onderliggende structuur duidelijk te maken. Dit wordt gedaan in de volgende paragraaf.

4.3.3. Stabiliteit van de factoren

De stabiliteit van de factoren is met behulp van drie split-half experimenten getest.

- a. De kwalitatieve interviews zijn verdeeld over twee groepen en daar is de PCA opnieuw uitgevoerd voor de waarden* per groep.
- c. De beoordelaars van de afstanden tussen de waarden* zijn random verdeeld over twee deelsteekproeven, waarna de PCA voor beide groepen opnieuw is uitgevoerd.
- d. De 160 waarden* zijn random in tweeën verdeeld (2 x 80) en ook daar is de PCA opnieuw per groep uitgevoerd.

Experiment a levert goed vergelijkbare factoroplossingen; dit wijst erop dat het aantal kwalitatieve interviews niet te gering is om stabiele uitkomsten te krijgen. Ook experiment b levert vergelijkbare factoroplossingen op; dit wijst erop dat de afstanden door een voldoende aantal proefpersonen zijn beoordeeld. Experiment c is de zwaarste en inhoudelijk de meest kritieke test omdat deze iets zegt over de vraag of de dimensies afhangen van de specifiek gekozen waarden*. Daarom wordt dit experiment uitvoeriger besproken.

De uitkomsten van beide deelsteekproeven worden beoordeeld op basis van de factorscores van de waarden. Het is onmogelijk om de beide deelsteekproeven direct te vergelijken, omdat elk van de waarden* slechts in één van de twee deelsteekproeven voorkomt en er dus geen vergelijkingsbasis voor factorscores is. Wel worden de deelsteekproeven vergeleken met de totale steekproef, omdat daarin wel factorscores van de waarden* bekend zijn. In tabel 3 wordt door middel van multiple correlaties aangegeven in welke mate de factoren uit de deelsteekproeven de factoren uit de totale steekproef kunnen verklaren. De correlaties zijn behoorlijk hoog, wat aangeeft dat de factoroplossing stabiel is. Dit resultaat is erg belangrijk, omdat het aangeeft, dat hoewel er geen garantie is voor het compleet zijn van de 160 waarden*, hun onderliggende betekenisstructuur stabiel is en kennelijk niet erg afhankelijk is van de gekozen omschrijving, zolang alle waardedimensies maar voldoende vertegenwoordigd zijn.

Tabel 3. (Multiple) correlaties factorscores tussen hele steekproef met 160 waarden en de twee deelsteekproeven met 80 waarden.

	Dimensie 1 160 waar- den	Dimensie 2 160 waar- den	Dimensie 3 160 waar- den	Dimensie 4 160 waar- den	Dimensie 5 160 waar- den	Dimensie 6 160 waar- den
Deelsteekproef 1 van 80 waarden	.972	.907	.796	.718	.826	.616
Deelsteekproef 2 van 80 waarden	.961	.788	.770	.900	.903	.879

5. TOEPASSING VAN DE INVENTARISATIE

Tot op heden wordt de SWOCC waarde-inventarisatie op twee manieren in de marketingwereld toegepast. Allereerst wordt de cd-rom gebruikt om ideeën te genereren, vooral de thesaurus wordt daarvoor gebruikt. Via deze thesaurus kunnen alternatieven worden gevonden voor waarden die weliswaar op veel producten van toepassing zijn, maar eigenlijk steeds iets anders bedoelen. Een voorbeeld hiervan is het in de markt zetten van producten die volgens de marketeer te maken hebben met gezelligheid. Het maakt echter nogal wat verschil of men spreekt over ‘gezelligheid’ bij een financiële instelling of ‘gezelligheid’ bij koffie of bij bier. Door in de thesaurus de waarde ‘gezelligheid’ op te zoeken en vervolgens te verbinden met de waardeomschrijvingen die aan ‘gezelligheid’ gekoppeld zijn, kan deze waarde bij de financiële instelling worden vervangen door ‘niet afstandelijk zijn’ en ‘je op je gemak voelen’; bij koffie door ‘gastvrij zijn’, ‘warmte’, sfeer’ of ‘je eigen spullen om je heen hebben’ en bij bier door ‘een feest geven’, ‘lol hebben’ of ‘mensen om je heen hebben’. Door de thesaurus worden deze mogelijkheden in een oogopslag duidelijk en kan de waarde ‘gezelligheid’ op een steeds andere manier worden verwoord. Ook geeft de cd-rom informatie over de verschillende manieren waarop mensen met totaal andere achtergronden over dezelfde items spreken.

Een tweede toepassing is het gebruik van de zes SWOCC-waardedimensies, met de daaraan gekoppelde 48 meest kenmerkende waarden, voor marktonderzoek. Hierbij worden respondenten verzocht aan te geven hoe belangrijk de 48 waarden in hun leven zijn en vervolgens of ze willen aangeven of ze die 48 waarden wel of niet van toepassing vinden op een bepaald merk of product. De WaardenMonitor die VNU in januari 2000 op de markt heeft gebracht is op deze dimensies en waarden gebaseerd. VNU heeft voor het ontwikkelen van de WaardenMonitor een groot aantal tijdschriften, lezers van die tijdschriften en een aantal merken onderzocht op waarden. Hierdoor is de mogelijkheid ontstaan om reclame voor de verschillende merken te kunnen plaatsen in het meest geschikte tijdschrift op basis van de waarden die respondenten aan de merken en tijdschriften hebben toegedeeld. Ook het communicatie-adviesbureau Senioragency[®] doet onderzoek onder ouderen met behulp van de SWOCC-waardedimensies zoals deze door VNU operationaliseerd zijn.

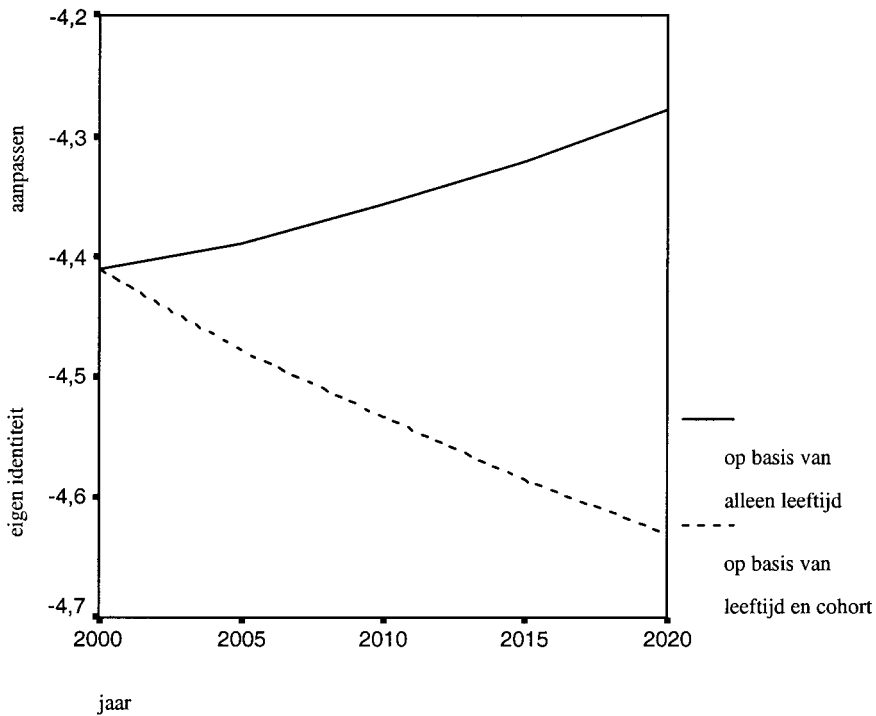
6. WAARDEN VAN DE TOEKOMST

Nu uit het onderzoek van VNU kan worden afgeleid welk belang mensen op dit moment aan bepaalde waarden toekennen, wordt het interessant om een blik in de toekomst te werpen. Immers voor het ontwikkelen van beleid moet vooruit gekeken worden. De vraag die hierbij aan de orde komt is: zijn waarden gebonden aan leeftijd of aan generatie? Indien blijkt dat de waardeoriëntatie alleen samenhangt met leeftijd kan men zich voor de toekomst richten op specifieke leeftijdsgroepen die min of meer constant blijven, maar als de waardeoriëntatie generatiegebonden is, vergt dat een geheel andere aanpak. Immers indien men zich dan op een bepaalde leeftijdsgroep richt, met een bij die groep behorende waardeoriëntatie, dan is dat slechts een tijdelijke zaak, omdat die specifieke groep na verloop van tijd verdwenen is en vervangen wordt door personen die een andere opvoeding hebben gehad en dus een andere waardeoriëntatie hebben.

Bij het analyseren van de resultaten, die beschikbaar zijn gekomen bij het ontwikkelen van de VNU WaardenMonitor, is gebleken dat er vooral verschillen bestaan tussen de waardeoriëntatie van personen van verschillende leeftijd en verschillende opleiding. Ook het Socioconsult onderzoek van Motivaction laat duidelijke verschillen zien in het belang dat personen van verschillende leeftijd aan waarden hechten (Spangenberg, Lampert & Mur, 1999; Spangenberg, Lampert & Peeck, 2000). Volgens Hofstede (1991) en Engel (1995) worden waarden voornamelijk in de jeugd aangeleerd, waarbij zowel het gezin als de school grote invloed hebben. Om te onderzoeken of ervaringen uit de jeugd invloed hebben op het waardepatroon op latere leeftijd is aan de leden van het CentERpanel een lijst voorgelegd met vragen over de jeugdervaringen (Hoe vaak ging u als kind naar de kerk? Mocht u vriendjes meenemen naar huis?) en een aantal uitspraken waarvan moest worden aangegeven of die een rol hebben gespeeld in hun opvoeding, zoals 'altijd met twee worden spreken', 'God ziet alles', 'geen papiertjes op de grond gooien', enzovoort. Het belang dat dit panel aan bepaalde waarden hechte was al bekend uit het onderzoek voor VNU.

Het rekening houden met generatiegebonden verschillen in jeugdervaring bleek cruciaal te zijn bij het maken van toekomstprojecties (Oppenhuisen, 2000). Een voorbeeld van zo'n projectie staat in figuur 4. Er worden twee prognoses gemaakt. Eén waarbij uitsluitend gekeken wordt naar de effecten van leeftijd, rekening houdend met de toenemende vergrijzing en één waarbij jeugdervaring een rol speelt. In de eerste variant wordt 'je aanpassen' in de toekomst belangrijker terwijl, als ook het generatie- of cohorteffect wordt verdisconteerd, het hebben van een 'eigen identiteit' juist van groter belang wordt.

Meer in het algemeen geldt dat als de resultaten op basis van alleen leeftijd worden bekeken, het lijkt alsof de waardeoriëntatie weer terugkeert naar de jaren 50. Een grote invloed van gezinswaarden, zorg voor anderen, veiligheid, aanpassen aan anderen en hedendaagse waarden als ontspannen zijn worden erg belangrijk. Wordt de jeugdervaring echter in de voorspelling meegenomen, dan is het resultaat een nog grotere nadruk op vrijheid, je eigen weg gaan, gericht zijn op jezelf, een uitdaging zoeken en waarden als stoer zijn, kortom de waarden die nu al zo'n grote rol spelen in onze samenleving worden alleen maar belangrijker. Voor het te ontwikkelen beleid, zowel in bedrijven als in de politiek, maken deze twee scenario's nog al wat verschil. Naarmate een volgende generatie tot de doelgroep gaat behoren, zou het beleid moeten worden bijgesteld, omdat de waarden van die groep niet dezelfde zijn als de waarden van de huidige doelgroep.



Figuur 4. De tegenstelling tussen je aanpassen en een eigen identiteit hebben.

7. CONCLUSIES

Dit onderzoek naar waarden bevat een methodologische en een inhoudelijke component. Het methodologische doel was het ontwerpen, uitvoeren en evalueren van een methode om waarden te selecteren. Deze methode moest herhaalbaar en controleerbaar zijn en in zekere zin compleet. Met de zorg die hieraan is besteed onderscheidt dit onderzoek zich van de meeste andere fundamentele waardeonderzoeken. De belangrijkste argumenten waarom dit onderzoek volledigheid kan claimen (uiteraard binnen de context van de Nederlandse cultuur en de beperking tot algemene waarden) zijn:

1. De 160 waarden blijken al na elf interviews min of meer compleet. Vervolg onderzoek heeft tot op heden slechts één nieuwe waarde kunnen toevoegen.
2. De gevonden factorstructuur blijft in stand bij halvering van het aantal waarden.

Daarnaast is dit onderzoek volledig herhaalbaar en controleerbaar, omdat de cd-rom een totaal overzicht geeft van alle gevonden resultaten en de genomen stappen om die resultaten te verkrijgen.

Het onderzoek heeft ook inhoudelijk resultaten opgeleverd. De 160 waarden worden gerepresenteerd door zes dimensies, die stabiel en robuust blijken voor wat betreft de gebruikte waarden. Deze dimensies kunnen worden geïnterpreteerd als bindingen aan zeer abstracte, maar wel zeer herkenbare levensdomeinen: relaties, sociaal georiënteerd zijn, het verleden, zekerheid, gezinsleven en conformistisch zijn. Binnen deze domeinen kunnen, volgens de Nederlandse respondent, waarden van elkaar verschillen voor verschillende mensen. En kennelijk zijn deze verschillen relevant.

Deze resultaten worden bevestigd door veel andere studies, zowel in een commerciële als in een wetenschappelijke context uitgevoerd. Een groot verschil met die andere onderzoeken is echter, dat dit onderzoek uitgaat van geheel nieuw basis materiaal, namelijk de waarden van de Nederlander. Een ander verschil is het aantal van zes dimensies in plaats van twee of drie. Toepassing van deze zes dimensies in marktonderzoek heeft al de winst van dit grotere aantal laten zien. Markten die op de ene dimensie niet scoren doen dat wel op andere dimensies, wat een verfijning van de te hanteren instrumenten toelaat.

De bruikbaarheid voor de marketing wereld kan worden vergroot door deze algemene waarde-inventarisatie uit te breiden met domeinspecifiek en product- of merkspecifiek onderzoek. Inmiddels zijn daar de eerste resultaten van bekend. Alles wijst erop dat de fundamentele waarden worden afgedekt door dit onderzoek, maar dat er inderdaad nog aparte domein- en productwaarden gevonden worden.

LITERATUUR

- Abramson, P. R., & Inglehart, R. (1995). *Value change in global Perspective*. Michigan: The University of Michigan Press.
- Becker, J. W., & Nauta, A. P. N. (1983). Enkele gegevens over waarden in Nederland na 1945. In J. W. e. a. Becker (Ed.), *Normen en waarden: verandering of verschuiving?* 's-Gravenhage: VUGA.
- Driel, H. v. (2000). *Het Waardenlabyrinth, waarden in Nederlandse soaps: fictie of reflexie?* (Doctoraalscriptie) Amsterdam: Universiteit van Amsterdam.
- Engel, J. F., Blackwell, R. D., & Miniard, P. W. (1995). *Consumer Behaviour*. (8 ed.). Orlando: The Dryden Press.
- Halman, L. (1991). *Waarden in de westers wereld. Een internationale exploratie van de waarden in de westerse samenleving*. Tilburg: University Press.
- Hofstede, G. (1991). *Cultures and Organisations, software of the mind*. Londen: McGraw-Hill Book Company.
- Kraemer, P. E. (1980). *Waardenonderzoek, wat is het waard?* Rotterdam: I.N.W. bulletin 24.
- Lautmann, R. (1971). *Wert und Norm: Begriffsanalysen für die Soziologie*. Opladen: Westdeutsche Verlag.
- Oppenhuisen, J. D. (2000). Een schaap in de bus? Een onderzoek naar waarden van de Nederlander. Amsterdam: SWOCC.
- Posthouwer, I. (1998). *De moraal van het verhaal. Waarden in allochtone en autochtone kinderboeken*. (Doctoraalscriptie) Amsterdam: Universiteit van Amsterdam.
- Reynolds, T. J., & Gutman, J. (1988). Laddering theory, method, analysis, and interpretation. *Journal of advertising research* (feb/mrt), 11-31.
- Rokeach, M. (1973). *The nature of human values*. New York: The Free Press.
- Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology*, 53(3), 550-562.
- Spangenberg, F., Lampert, M. A., & Mur, J. P. (1999). *De maatschappelijke betrokkenheid van jongeren. Een socioconsult onderzoek voor de Vereniging voor Bestuurskunde*. Amsterdam: Motivaction.
- Spangenberg, F., Lampert, M. A., & Peeck, V. H. (2000). *Socioconsult-onderzoek naar woonwensen van senioren en toekomstige senioren. 'Een goede buur'*. Amsterdam: Motivaction.
- Vinson, D. E., Scott, J. E., & Lamont, L. M. (1977). The role of personal values in marketing and consumer behavior. Can personal values be used to assist marketers in determining consumer choice behavior? *Journal of Marketing* (april), 44-50.

Bijlage 1: foto's en vragen.

Foto	Onderwerp	Vragen bij de foto(s)
1	Kinderen	Waarom nemen mensen kinderen waarom willen mensen geen kinderen waarom willen mensen een groot gezin
2	School	wat betekent school voor mensen/ wat heeft school voor u betekend waarom sturen ouders hun kinderen naar school → wat verwachten ze ervan
3	Lezen/ boeken	waarom lezen mensen boeken waarom willen mensen boeken hebben
4	Puberteit	waarom zien deze mensen er anders uit leef je eens in in deze personen → wat zal voor deze personen belangrijk zijn in hun leven (waarden)
5	Huwelijk/ liefde/ seks	waarom willen mensen trouwen waarom willen mensen trouwen met alles erop en eraan waarom willen ze de familie erbij hebben waarom kan het stuk lopen waarom willen mensen met elkaar spelen waarom willen mensen met elkaar vrijen
6	Dieren	waarom hebben mensen dieren verplaats je eens in deze mevrouw, wat zal voor deze mevrouw belangrijk zijn in haar leven
7	Vriendschap	waarom hebben mensen vrienden waarom zoeken veel mensen de drukte op
8	Werk	waarom werken mensen waarom willen mensen geld verdienen waarom willen mensen hogerop komen leef je eens in in deze personen, wat zal voor hun belangrijk zijn
9	Huis	waarom willen mensen een eigen huis hebben waarom zien alle huizen er zo verschillend uit, zowel van buiten als van binnen
10	Vakantie/ sport/ gezondheid	waarom gaan mensen op vakantie waarom willen ze uitrusten waarom willen ze spanning waarom sporten mensen waarom is gezondheid zo belangrijk
11	Arm/ rijk	welke tegenstelling zit hier in welke waarden zullen voor de twee groepen belang- rijk zijn
12	Geloof	waarom gaan mensen naar de kerk waarom geloven mensen waar draait het in het leven om
13	Koningshuis/ leger	waarom willen mensen de koningin/ koningshuis waarom willen ze het niet waarom hebben we een leger in WO II waren er mensen in het verzet actief. Wat voor kenmerken moeten deze mensen gehad hebben

		om dat te kunnen doen. waarom willen mensen in het leger waarom willen mensen niet in het leger waarom zal een vrouw het leger in willen geef eens een beschrijving van de Nederlandse bevolking
14	Technologie/ toekomst	waarom is de mens met technologie bezig waarom betekent de toekomst voor mensen
15	Milieu	wat zie je op deze foto's wat zijn de verschillen tussen de foto's
16	Uiterlijk	waarom gebruiken mensen schoonheidsmiddelen waarom is uiterlijk belangrijk voor de mens
17	Eten/ snoepen	waarom eten mensen soms een hele zak koek of drop leeg
18	Oud/ jong	waarom willen mensen jong blijven waarom is in andere culturen juist de oudere mens meer in tel
19	Dood	wat betekent doodgaan voor mensen waarom zijn veel mensen bang van de dood
20	Zorgen voor anderen	waarom zorgen mensen voor andere mensen hoe vinden mensen het dat er voor hen gezorgd wordt
21	Geweld	waarom hebben we politie waarom gebruiken mensen geweld
22	Politiek	waarom willen mensen in de politiek waarom zijn dan oudere mensen opeens wel weer in tel

2. *Kwaliteitszorg en professionalisering in het marktonderzoek*

A.E. BRONNER en T.J. H. VELDMAN

SAMENVATTING

In de afgelopen 25 jaar zijn er instrumenten ontwikkeld die de basis vormen voor het leveren van kwaliteit door marktonderzoekbureaus. Te denken valt aan minimumeisen voor veldwerk, ESO-MAR-code, VMO kwaliteitseisen, Total Quality Management en ISO-certificering. Toch lijken de marktonderzoekbureaus er niet altijd in te slagen de markt voldoende duidelijk te maken wat zij aan kwaliteitszorg doen. Dat heeft ons ertoe gebracht een systematisch overzicht voor dit Jaarboek op te stellen met betrekking tot de inspanningen die in de loop van de jaren zijn verricht. Hoe heeft de discussie over de kwaliteitszorg zich in de laatste decennia ontwikkeld en welke overwegingen hebben daarbij een rol gespeeld?

En naar de toekomst kijkend: hoe zou die kwaliteitszorg eruit moeten gaan zien en welke overwegingen spelen daarbij? Misschien Michelin-sterren voor onderzoeksbureaus?

1. INLEIDING

In de praktijk van het marktonderzoek hebben we te maken met opdrachtgevers die opdrachten tot onderzoek verstrekken en marktonderzoekbureaus die deze projecten uitvoeren. In interviews in de vakpers met opdrachtgevers kan men regelmatig uitspraken tegenkomen die aangeven dat kwaliteit en kwaliteitszorg in het marktonderzoek heel belangrijk door hen worden gevonden.

Talrijk zijn de deskundigen die zich hebben gebogen over een beschrijving van het begrip kwaliteit en even talrijk zijn daarmee ook de verschillen duidelijk geworden die aan de opvattingen over kwaliteit kunnen worden gekoppeld. Het ligt voor de hand dat ook een definitie van kwaliteit in het marktonderzoek niet eenduidig is. En in een markt waar zoveel bureaus van verschillende pluimage en omvang werkzaam zijn telen uiteraard de eigen specifieke opvattingen over kwaliteit welig. Maar daarnaast bestaat er binnen een groot deel van de beroepsgroep, verenigd in de voormalige Vereniging van Marktonderzoekbureaus, een nadrukkelijke behoefte opdrachtgevers duidelijk te maken dat zij kunnen rekenen op een door hen allen gedeelde opvatting over de kwaliteit van de dienstverlening. Het zoeken is dus naar de gemeenschappelijke deler in de opvattingen over kwaliteit. In de afgelopen jaren zijn er instrumenten ontwikkeld en toegepast die de basis vormen voor het leveren van kwaliteit door marktonderzoekbureaus, zoals minimumeisen voor veldwerk, kwaliteitseisen voor kwantitatief en kwalitatief onderzoek, Total Quality Management en ISO-certificering.

Toch lijken de marktonderzoekbureaus er niet altijd in te slagen de markt voldoende duidelijk te maken wat zij aan kwaliteitszorg doen. Uit het brancheonderzoek dat Motivaction elke twee jaar uitvoert blijkt volgens de laatste meting in oktober/november 2000 (Motivaction 2001) dat voor lang niet alle opdrachtgevers de ISO-certificering een aanbeveling is om met het betreffende bureau te werken. Dit enigszins magere resultaat heeft ons ertoe gebracht een systematisch overzicht voor dit Jaarboek op te stellen met betrekking tot de inspanningen die in de loop van de jaren zijn verricht. Hoe heeft zich de discussie over de kwaliteitszorg in de laatste decennia's ontwikkeld, welke overwegingen spelen daarbij een rol en hoe staat het marktonderzoek er in dit opzicht nu voor? We zullen niet alleen naar het verleden kijken, maar ook naar de toekomst.

2. DE EERSTE STAP: MINIMUMEISEN VELDWERK

De eerste stukken over kwaliteitszorg door marktonderzoekbureaus werden eind jaren zeventig opgesteld door de Werkgroep Buro Overleg (WBO). Men richtte zich vooral op het veldwerk, en het eerste stuk heette dan ook 'Minimum eisen ten aanzien van controle op uitgevoerd veldwerk'. Men besloot als regel voor te schrijven dat een controle bij 20% van de gerealiseerde vraaggesprekken uitgevoerd moest worden. Bij bepaalde verdenkingen zou het volledige werk van de betreffende enquêteur/enquêtrice gecontroleerd moeten worden en "wel zodanig dat een uiteindelijk goed oordeel gevormd kan worden". Voorts werd afgesproken dat alle WBO-bureaus een administratie zouden gaan bijhouden waarin per project een verslag van de uitgevoerde controle werd gegeven. Daarna werden nog in het begin van de jaren tachtig minimumeisen voor steekproef- en non-respons verantwoording toegevoegd. Een verdere uitbreiding van de minimumeisen werd door sommige leden van de WBO wel nagestreefd, maar zoals Emile van Westerhoven, die op dit terrein zeer actief was, aangeeft: "Het was niet mogelijk alle bureaus op één lijn te krijgen."

Veel energie werd ook gestoken in de opzet van een zelfregulerend orgaan waarbij opdrachtgevers financiële klachten en kwaliteitsklachten zouden kunnen indienen. Uit een notitie uit die tijd citeren wij (van Westerhoven, 1984, pg.4):

"Het werkterrein van een zelfregulerend orgaan zal dus betrekking hebben op de kwaliteit van de geleverde prestatie. Nu is kwaliteit een rekbaar begrip. Het hanteren van normen door dogmatici kan leiden tot het door de mand vallen van welhaast ieder onderzoek, terwijl aan de andere kant het laten prevaleren van praktische overwegingen teveel aan de inhoudelijke kwaliteit van onderzoek voorbijgaat. Willen we kwaliteit ten behoeve van een zelfregulerend orgaan definiëren, dan moet naar het oordeel van de commissie in eerste instantie worden uitgegaan van de afspraken, die tevoren zijn gemaakt, in de vorm van een voorstel tot onderzoek – projectspecificatie, opdrachtbrief, etc. Extreem gezien houdt dit in, dat opdrachtgevers die 'slecht' onderzoek bestellen, achteraf minder kwaliteit kunnen claimen dan opdrachtgevers, die kwalitatief hoogwaardig onderzoek bestellen."

Uit het citaat blijkt dat men niet probeerde min of meer 'objectieve' kwaliteitsregels te formuleren, maar meer een toets op gemaakte afspraken nastreefde. Maar uiteindelijk bleef het zelfregulerend orgaan in intenties steken.

Samenvattend kunnen we stellen dat de WBO de eerste aandacht in Nederland voor gezamenlijke kwaliteitsregels stimuleerde. Maar dat dat eigenlijk alleen maar lukte wat betreft direct aan het veldwerk gelieerde regels.

Midden jaren tachtig schreef van Westerhoven: “De WBO lijkt onmachtig om als geschikt platform te dienen voor gezamenlijke stappen ter kwaliteitsverhoging, deels omdat besluiten in de hier aangeduide sfeer bij unanimiteit moeten worden genomen, deels ook omdat gemelde onderlinge tegenwerking zich binnen de WBO lijkt te demonstreren.”

Dat wil overigens niet zeggen dat het marktonderzoek geheel zonder regelgeving fungeerde. Kwaliteitsaspecten komen ook naar voren in de afbakening van rechten en plichten van onderzoekers, opdrachtgevers, respondenten en overheid. Een voorbeeld hiervan zijn de regels van de European Society for Opinion and Marketing Research (ESOMAR). De eerste professionele gedragscode die de manier regelt waarop marktonderzoekprojecten worden uitgevoerd stamt uit 1948. Deze werd gevolgd door andere organen, zoals de Internationale Kamer van Koophandel (ICC). In 1976 besloten ESOMAR en ICC tot een gezamenlijke internationale code die voor het eerst in 1977 werd gepubliceerd.

Mede naar aanleiding van de totstandkoming van de Wet Persoonsregistraties (1988) namen de Nederlandse Vereniging van Marktonderzoek (NVvM) en de WBO in de loop van 1990 het initiatief tot het instellen van een klachtencommissie die “belast wordt met het toezicht op de naleving van de Gedragscode met betrekking tot Marktonderzoek, Persoonsgegevens en Privacybescherming en met het beslechten van geschillen tussen respondenten enerzijds en/of marktonderzoekbureaus anderzijds, die bij de toepassing van deze gedragscode rijzen”. In de praktijk is gebleken dat de bemoeienissen van dit College van Klachtenbehandelaars zich niet hebben beperkt tot bemiddeling in geschillen met betrekking tot de WPR, maar op verzoek van opdrachtgevers ook tot uitspraken over de kwaliteit van uitgevoerde onderzoeksprojecten.

3. DE TWEEDE STAP: REGELS EN RANDVOORWAARDEN

De opvolger van de WBO was de VMO (Vereniging van Marktonderzoekbureaus), een bredere organisatie waartoe bureaus gemakkelijker toegelaten konden worden en wat minder een ‘herenclub’ waarin de oprichters en stimulators van het marktonderzoek onder een goede maaltijd zaken regelden. Dat bood ook weer nieuwe perspectieven voor de kwaliteitsregels.

Fred Bronner en Rob van Leeuwen gingen in 1988 aan de slag en namen de wijze raadgevingen van Emile van Westerhoven ter harte. De VMO Commissie Kwaliteit besloot de unanimiteit los te laten en koos voor de volgende procedures:

- opstelling van een uitgebreide lijst met kwaliteitscriteria
- per punt dienden de VMO-leden aan te geven op een vijfpuntsschaal wat zij ervan vonden
 - zeker opnemen
 - wel mee eens
 - punt van discussie
 - niet mee eens, maar als de meerderheid wil wat mij betreft o.k.
 - helemaal niet mee eens, zeker niet opnemen

- op basis van de enquête-uitslag (20 bureaus) werd een voorstel gedaan tot wel/niet opnemen van een kwaliteitseis
- plenaire discussie
- afronding.

In juni 1991 presenteerde de VMO de ‘VMO minimum kwaliteitseisen voor kwantitatief en kwalitatief onderzoek’. Opvallend was ten opzichte van 10 jaar daarvoor de consensus onder de onderzoeksbureaus (zie ook Kuin, 1991, p. 34).

“De eisen aan het kwantitatieve onderzoek zijn verdeeld in ‘items’ die standaard in het rapport moeten worden opgenomen, en die niet standaard zijn maar na overleg toch aan de opdrachtgever zijn te leveren. Een greep uit de eerste groep. Duidelijk moet zijn of er wel of geen onderaannemers zijn ingeschakeld voor het veldwerk. En of er al dan niet projecten zijn gecombineerd in één interview. De opdrachtgever hoort de vragenlijst te kunnen krijgen. De onderzoeksmethode hoort te zijn omschreven: face to face, schriftelijk, telefonisch of een combinatie. Hetzelfde geldt voor het onderzoeksdoel en de probleemstelling. Verder moet er duidelijkheid zijn over onder meer de populatie, de steekproeftrekking, clustervorming en de interviewperiode, de controlemethode en de weging. Belangrijk is ook het voorschrift dat het onderzoeksrapport een duidelijk onderscheid maakt tussen beschrijvingen aan de ene kant en speculaties plus veronderstellingen aan de andere kant. Niet standaard maar toch mogelijk zijn bijvoorbeeld het aangeven van statistische significantie van verschillen en het noemen van de gebruikte opties bij multivariate analyses. Voor zowel de voorstellen als de rapporten inzake kwalitatieve onderzoeken staan de volgende eisen op papier. Doelstelling en probleemstelling moeten zijn omschreven. De respondentenselectie moet worden verantwoord. Duidelijk moet zijn of onderzoeksdeelnemers al dan niet ‘verhuld’ zijn uitgenodigd. De groepsgrootte moet worden vermeld. Hetzelfde geldt voor de interviewperiode, de gespreksduur en de gespreksplaats.”

Opdrachtgevers reageerden wat kritisch omdat zij niet waren betrokken bij de opstelling van de regels, maar waren toch verheugd, want nu hadden ze tenminste ‘een hulpmiddel, een checklist waar de VMO voor staat, en waar de club van onderzoeksbureaus op kan worden aangesproken’.

De VMO-kwaliteitseisen waren een goed startpunt en in overleg tussen de Commissie Kwaliteit (later versterkt met Jan Urlings) en de Commissie Toetsing en Toelating vonden op basis van de praktijk updates plaats (1994 en 1996).

De Commissie Toetsing en Toelating visiteerde de bureaus met enige regelmaat en controleerde of de bureaus zich aan de kwaliteitsrandvoorwaarden hielden. Deze commissie memoreerde over deze eisen in 1994 het volgende.

“In de promotion door de VMO en haar leden wordt nogal eens de nadruk gelegd op de kwaliteitseisen die worden gehanteerd. Daarmee zouden zij een zodanige garantie voor de opdrachtgever vormen dat zij op ‘kwaliteitsonderzoek’ kunnen rekenen. De waarde van deze opvatting is evenwel betrekkelijk. Niet dat de VMO-bureaus geen kwaliteit zouden leveren, maar wel dat de kwaliteitseisen daarvoor een zekerheid zouden bieden. Want wat houden die kwaliteitseisen nu eigenlijk in? In essentie is het niet meer dan een opsomming van voorwaarden/elementen die in een rapport of voorstel moeten worden vermeld. De controle daarop heeft vooral een formeel karakter op de

aanwezigheid van de vereiste elementen, maar niet op een inhoudelijke kwaliteit” (Veldman, 1994).

Maar ondanks de diversiteit aan opvattingen overheerste in de periode 1991 – 1994 optimisme omdat de bureaus voor het eerst een uitgebreid samenstel aan randvoorwaarden gezamenlijk onderschreven.

4. PROCESKWALITEIT: ISO 9001

In een algemene ledenvergadering van de VMO in oktober 1993 werd expliciet als onderdeel van het ‘mission statement’ aangenomen “het bevorderen van de kwaliteit en de professionalisering van het vak”. En dat impliceerde het startschot voor het gaan werken aan ISO-certificatie. Er werd contact gezocht met CERTIKED, een organisatie die onderzoek had gedaan naar de mogelijkheid van certificatie voor ondernemingen en organisaties die zich bewegen op het terrein van de kennisintensieve dienstverlening. Toenemende internationalisering en eisen vanuit de overheid brachten de bureaus op het ISO-spoor. Om een indruk te geven van het soort eisen dat in deze ISO-regelgeving wordt vastgelegd, volgt hier een korte opsomming van enkele cruciale punten:

- de leverancier definieert en documenteert zijn beleid en doelstellingen en garantie met betrekking tot kwaliteit
- een gedocumenteerd kwaliteitssysteem wordt opgesteld met het doel ervoor te zorgen dat de geleverde diensten aan specifieke eisen voldoen
- verantwoordelijkheden, bevoegdheden en de interrelaties tussen alle personeelsleden die de werkzaamheden leiden, uitvoeren en controleren behoren te worden vastgelegd, in het bijzonder voor degenen die organisatorische vrijheid en bevoegdheid behoeven te aanzien van kwaliteitsbewaking
- de nodige controlemaatregelen worden vastgelegd, er wordt gezorgd voor de nodige middelen en voor de opleiding van degenen die met de controle belast zijn
- een kwaliteitsmanager wordt aangesteld.

De VMO stelde in 1994 een handleiding voor een kwaliteitshandboek op, die de leden konden benutten bij het vorm geven van een eigen kwaliteitshandboek.

Over kwaliteit werd gesteld dat het een relatief begrip is dat pas betekenis krijgt als de bijbehorende eisen bekend zijn. Het is in feite de balans tussen: enerzijds de eisen/wensen/ uitgesproken en niet-uitgesproken verlangens, anderzijds de eigenschappen/kenmerken/ resultaten van een product, proces of dienst.

In een VMO-stuk (VMO, 1994) wordt aangegeven dat drie begrippen centraal staan:

- (a) kwaliteitssysteem: de organisatorische structuur, verantwoordelijkheden, procedures, processen en voorzieningen voor het ten uitvoer brengen van kwaliteitszorg
- (b) kwaliteitsbeheersing: de operationele technieken en activiteiten die worden toegepast om te bewerkstelligen dat aan de kwaliteitseisen wordt voldaan
- (c) kwaliteitsborging: het geheel van alle geplande en systematische acties nodig om in voldoende mate het vertrouwen te geven dat een product of dienst voldoet aan de gestelde kwaliteitseisen.

Aan de leden van de VMO werd tot 2000 de tijd gegeven om het ISO-9001 certificaat te behalen.

Vanaf 1995 voldeden steeds meer bureaus aan de ISO-normen. Door het ISO-certificaat te combineren met het aantoonbaar rekening houden met de VMO-kwaliteitseisen (zie de vorige paragraaf) ontstond de aanduiding ISO-plus.

Alle VMO-bureaus zouden op deze wijze tussen 1995 en 2000 ISO-plus moeten worden. De ISO-regels garandeerden een kwalitatief hoogwaardig proces (proceskwaliteit), de VMO-kwaliteitsregels garandeerden een kwalitatief hoogwaardige output (productkwaliteit).

Marktonderzoekbureaus hebben aanzienlijke investeringen gedaan om tot ISO-certificering te komen in de verwachting dat dat bij opdrachtgevers een duidelijke hogere kwaliteitsperceptie van de marktonderzoekbranche zou opleveren. Maar de opdrachtgevers blijken dat niet of anders te percipiëren. Elke twee jaar voert Motivaction een brancheonderzoek uit onder opdrachtgevers met enerzijds vragen over het marktonderzoek in het algemeen en anderzijds vragen over het imago van specifieke bureaus. De laatste meting werd in januari 2001 gepubliceerd en het veldwerk vond plaats in oktober/november 2000 (Motivaction, 2001). Hieruit blijkt dat een ISO-certificering voor ongeveer een derde van de opdrachtgevers een aanbeveling is om met het betreffende bureau te werken. Dat is opvallend laag, maar nog opvallender is dat ongeveer een kwart het hebben van een ISO-certificaat *juist niet* als een aanbeveling ziet. De groep respondenten die ISO-certificering niet als een aanbeveling ziet om met een bureau te werken is significant groter dan in 1998; 10% meer in de totale steekproef.

Ongetwijfeld heeft certificering geleid tot een zodanige organisatie van een bureau dat verantwoordelijkheden voor iedereen duidelijk zijn en dat er minder fouten worden gemaakt. Maar Veldman constateerde in 1997 dat in de door hem onderzochte kwaliteitshandboeken van acht bureaus de definiëring van het begrip kwaliteit niet veel anders opleverde dan 'het leveren van diensten overeenkomstig de eisen, wensen en behoeften van opdrachtgevers' of 'het verbeteren van de kwaliteit van de dienstverlening' of 'het leveren van een kwaliteitsniveau nodig om te voldoen aan de eisen van de klant' (Veldman, 1997).

5. VAN ISO NAAR TQM

In 1998 schreef de directeur van de VMO Görtemöller (Görtemöller, 1998) een notitie onder de titel 'Van ISO 9001 naar Total Quality Management'. Hij zag als een vervolgstap van ISO-plus de invoering van TQM. Total Quality Management, ook wel integrale kwaliteitszorg genoemd, is een benadering die door het herstructureren van traditionele managementpraktijken streeft naar het verbeteren van kwaliteit, het verlagen van de kosten en verbeteren van de tevredenheid van de afnemer (kwaliteit van output van bureaus).

Een belangrijk hulpmiddel om in ondernemingen over een breed front en integraal aan kwaliteitsverbetering te werken komt van de European Foundaton for Quality Management (EFQM).

Het EFQM model (zie figuur 1) begint bij kwalitatief leiderschap in de onderneming en mondt uiteindelijk uit in de kwalitatieve beheersing van de output.

Het model bestaat uit drie delen:

het **eerste deel** betreft de verschillende facetten in het ondernemingsbeleid waaraan gewerkt wordt om integrale kwaliteit te bereiken. Het betreft voornamelijk de maatregelen die een bedrijf neemt op het vlak van processen, personeel en middelen.

Leiderschap

Een consequent op IKZ (Interne Kwaliteitszorg) gerichte cultuur, tijdig erkenning en waardering tonen voor inspanningen en successen van individuen en teams; ondersteuning van IKZ door de juiste middelen en hulp beschikbaar te stellen; betrekkingen met klanten en afnemers; actieve bevordering van IKZ buiten de eigen organisatie.

Beleid en strategie

Hoe beleid en strategie op IKZ worden gebaseerd; hoe beleid en strategie worden opgesteld op basis van voor IKZ relevante informatie; hoe beleid en strategie worden gecommuniceerd; hoe beleid en strategie regelmatig worden getoetst en verbeterd.

Personeelsmanagement

Hoe continue verbetering van het personeelsmanagement wordt bereikt; hoe de vaardigheden en de capaciteiten van de medewerkers op peil worden gehouden en ontwikkeld door middel van werving, training en promotie; hoe individuen en teams in overleg doelen stellen en hun prestaties continu beoordelen; hoe ieders betrokkenheid bij de continue verbetering wordt bevorderd en mensen de bevoegdheid krijgen voor de juiste actie; hoe effectieve communicatie van boven naar beneden en omgekeerd wordt bereikt.

Middelen

Hoe wendt het bedrijf financiële middelen aan; hoe zet het informatiesystemen in; hoe wendt het materiële middelen aan; hoe springt het bedrijf om met technologie?

Management van de processen

Hoe worden de processen geïdentificeerd die kritisch zijn voor de onderneming; hoe beheert het bedrijf processen systematisch; hoe gebruikt het procesprestatie metingen en gegevens uit terugkoppeling om processen te beoordelen en verbeteringsdoelen te stellen; hoe stimuleert het innovatie en creativiteit bij procesverbetering; hoe voert het procesveranderingen in en evalueert het de voordelen ervan?

Het **tweede deel** handelt over het meten van de resultaten van de acties beschreven in deel één:

Klanttevredenheid

Wat vinden externe klanten van de kwaliteit van de geleverde diensten en het functioneren van de organisatie? (eigen meting)

Tevredenheid van de medewerkers

Wat vinden de medewerkers van hun organisatie? (interne meting)

Ondernemingsresultaten

Wat heeft de onderneming bereikt ten opzichte van de geplande kwaliteit van de output en andere maatstaven (ook financieel).

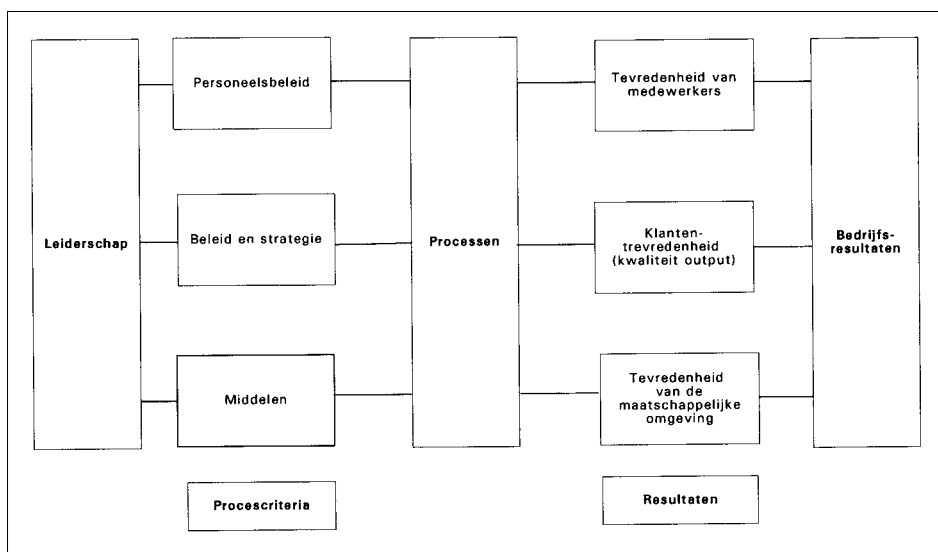
Het *derde deel* bestaat uit een onafhankelijk benchmarking onderzoek waarbij de volgende aspecten centraal zullen staan:

Klanttevredenheid

Hierbij zal de geleverde kwaliteit (kwaliteitsbeleving) een belangrijk onderdeel vormen.

Invloed op de omgeving

Wat vinden relevante doelgroepen van de branche. Dit betreft bijvoorbeeld ook meningen over de wijze waarop de ondernemingen omgaan met respondenten.



Figuur 1. Visualisatie van het EFQM-model (Görtemöller, 1998).

Uiteindelijk werd besloten aan de individuele leden over te laten in hoeverre men TQM ging implementeren binnen de organisatie.

6. VAN RANDVOORWAARDEN NAAR INHOUDELIJKE BEOORDELING

Met het opgaan van de VMO in de MOA (MarktOnderzoek Associatie) in 2000 werden de kwaliteitsregels opnieuw fundamenteel ter discussie gesteld. Er werd gesteld dat de bijna 10 jaar geldige VMO-kwaliteitsregels eigenlijk niet meer dan randvoorwaarden waren en vooral een functie hadden om bureaus duidelijk te laten vertellen wat zij precies allemaal gedaan hadden. De regels boden geen mogelijkheid aan te geven of een gekozen methode van dataverzameling wel de juiste was geweest en of de gekozen steekproefomvang wel in orde was. Er was behoefte om naast de formele eisen een meer inhoudelijke beoordeling van voorstellen en rapporten te geven. Wellicht zelfs

uitmondend in een soort sterrenstelsel voor onderzoeksbureaus (3 sterren betekent bv. dat alle gecontroleerde voorstellen en rapporten aan de formele eisen voldoen en relatief hoog scoren op de inhoudelijke beoordeling).

Deze behoefte aan meer inhoudelijke beoordeling werd gevoed door een drietal ontwikkelingen. Allereerst een in 1999 nog binnen VMO-verband uitgevoerd experiment. Het experiment werd uitgevoerd door leden van de Commissie Toetsing en Toelating naar de inhoudelijke kwaliteit van de werkzaamheden van twee bureaus. Per type onderzoek (b-to-b, kwalitatief- en kwantitatief onderzoek) werden vier uitgevoerde onderzoeksprojecten beoordeeld op hun inhoud. Daarbij vond een toetsing plaats op de volgende onderdelen:

Uitgaande van de probleemstelling:

- formulering van de doelstelling van het onderzoek
- keuze van de onderzoeksmethodiek
- volledigheid en relevantie van de vragenlijst

De uitvoering van het onderzoek in technische zin (aantal gesprekken, steekproef etc.)

Uitgaande van de probleemstelling:

- volledigheid en juistheid van de rapportage
- formulering conclusies resp. inhoud van de presentatie

Aan elk van de zes onderdelen werd een score van 'zeer goed' tot 'zeer onvoldoende' toegekend. Daarnaast werd nog rekening gehouden met enkele aspecten die een voorwaarde vormen om tot een goede kwaliteit te komen, zoals de wijze van organisatie, ISO-certificering, ervaring van projectleiders etc. Met dit experiment kon worden aangetoond dat metingen naar de inhoudelijke kwaliteit van onderzoek wel degelijk mogelijk waren.

Een volgende aanleiding tot een bezinning over de geleverde kwaliteit van onderzoek vormde het onderzoek van de afdeling Methoden & Technieken van de Faculteit Sociaal Culturele Wetenschappen van de Vrije Universiteit (H. van den Berg et al, 2000). Met name de soms onzorgvuldige wijze waarop de bureaus de zo populaire stellingvragen in de dagelijkse praktijk van het markt- en opinie-onderzoek toepassen wordt hier aan de orde gesteld.

Een derde reden om aandacht te schenken aan inhoudelijke kwaliteit is het samengaan van de NVMI en de VMO, waarbij de VMO-bureaus een doorstart maken onder de naam Research Keurmerkgroep. De term 'keurmerk' is bewust gekozen om kwaliteit in de uitvoering en de professionaliteit uit te kunnen dragen.

Deze roep om een meer inhoudelijke beoordeling is des te opmerkelijker als we naar ontwikkelingen in andere Europese landen kijken. ESOMAR, EFAMRO en individuele brancheorganisaties kwamen in 1998/1999 juist met publicaties en richtlijnen die sterk lijken op de VMO-kwaliteitsregels eerste versie uit 1991. (ESOMAR News Brief november 1998, EFAMRO Market Research Quality Standards 1998, Standards zur Qualitätssicherung in der Markt- und Sozialforschung oktober 1999). Nederland lijkt

dus in het streven naar een meer inhoudelijke beoordeling voorop te lopen. Reden om op deze discussie die momenteel binnen de Research Keurmerkgroep van de MOA speelt (tussen zomer 2000 en zomer 2001) wat dieper in te gaan.

7. DE INVULLING VAN EEN MEER INHOUDELIJKE KWALITEITSBEOORDELING

Zoals in de vorige paragraaf al is aangestipt, is de discussie momenteel (2001) om een splitsing aan te brengen tussen de formele eisen, die in feite een soort update zijn van de oude 1991 VMO-kwaliteits-regels en inhoudelijke beoordeling. In het begin van 2001 hebben de leden van de Research Keurmerkgroep als uitgangspunt gekozen een hiërarchie (zie figuur 2) van regels en procedures waar zo goed mogelijk recht wordt gedaan aan de wens om kwaliteit te leveren.



Figuur 2. Hiërarchie van Keurmerkeisen van de Research Keurmerkgroep.

De basis (niveau 1) vormt niet alleen afspraken over een minimumniveau qua organisatie en structuur van de organisatie, maar ook de reeks van besluiten die door de groep genomen zijn en die met name de wettelijke regelgeving volgen. De formele eisen die aan de inhoud van voorstellen en rapporten zijn gesteld hebben wat modernisering ondergaan op basis van de praktijk van het onderzoek, zoals de mogelijkheid een black box binnen een model niet geheel openbaar te maken, niet altijd meer een vragenlijst in het rapport, maar wel inzagerecht en sheet-rapportage met alleen conclusies op hoofdpunten. Kortom, een aantal aanpassingen die het gedachtegoed van onderzoeksbureaus beter beschermen. Ook omdat het met enige regelmaat voorkomt dat niet-onderzoeksbureaus ontwikkelde methoden en technieken van onderzoeksbureaus overnemen en in feite plagiaat plegen.

De inhoudelijke beoordeling kan worden onderscheiden in twee trappen (niveau 2 en niveau 3 in figuur 2). Allereerst (niveau 2) is er een beoordeling naar niveau van de organisatie en naar de gebruikswaarde. Gesteld is dat het niveau van de *organisatie* de basis vormt voor het leveren van een goede kwaliteit. Daarbij gaat het niet alleen om

opleidingsniveau, ervaring, training ed. van medewerkers, maar ook om de bijdrage die de organisatie levert aan het niveau van het marktonderzoek in het algemeen (artikelen, lezingen etc.). De *gebruikswaarde* kan worden omschreven als de mate waarin het onderzoek heeft bijgedragen tot het oplossen van het gestelde probleem in de ervaring van de opdrachtgever. De informatie over de gebruikswaarde moet worden gestandaardiseerd en zou als verplicht item in klanttevredenheidsonderzoeken kunnen worden opgenomen. Deze omschrijving van gebruikswaarde onderscheidt zich duidelijk van belevingsfactoren, zoals de wijze van omgang tussen opdrachtgever en onderzoeker, snelheid van levering van gegevens, het bureau-imago etc. Deze belevingsfactoren vormen geen onderdeel van de beoordeling van kwaliteit.

Er is voor gekozen zowel de definities van gebruikswaarde als die van niveau van de organisatie zo goed mogelijk meetbaar te maken, waardoor een vergelijking tussen bureaus mogelijk wordt.

De belangrijkste trap (niveau 3) in deze ladder van kwaliteit is zonder meer de beoordeling van (a) voorstellen en (b) onderzoeksopzet en –uitvoering.

Bij een inhoudelijke beoordeling van voorstellen zijn twee dimensies van belang:

1. de keuze van de onderzoeksmethodiek in relatie tot de onderzoeksvraag
2. de toegankelijkheid voor de lezer in termen van volledigheid, overzichtelijkheid en leesbaarheid.

Bij onderzoeksopzet en –uitvoering is kernpunt een ‘toetsing’ – in volgorde van importantie – op drie onderdelen:

1. de inhoud en toepassing van de vragenlijst, dat wil zeggen de keuze van de vragen in relatie tot de geformuleerde doelstelling van het onderzoek, de formulering van de vragen zelf en de wijze van interpretatie van de verkregen antwoorden
2. de formulering van de doelstelling van het onderzoek, gezien de probleemstelling en de wijze waarop verkregen onderzoeksgegevens aansluiten op de probleemstelling
3. de keuze van de onderzoeksmethodiek en de uitvoering van het onderzoek in technische zin.

Het valt niet te ontkennen dat met de invoering van dit toetsingssysteem subjectieve oordelen worden geïntroduceerd, waarover zonder twijfel grondige discussies kunnen ontstaan. Aan de andere kant raken we met deze criteria wel de kern van het markt- en opinieonderzoek. Teneinde toch een zo goed mogelijk oordeel over inhoudelijke kwaliteit op deze items te krijgen zal er voor worden gekozen die controles te laten uitvoeren door een gezelschap van deskundigen die geen binding (meer) hebben met specifieke bureaus en waarbij een ruime verscheidenheid van uitgevoerd onderzoek in de toetsing wordt betrokken.

De commissie die de controles uitvoert zal periodiek een rapportage publiceren over de kwaliteit van het onderzoek dat de leden van de Research Keurmerkgroep over een bepaalde periode hebben uitgevoerd. Op deze wijze krijgt ook de buitenwereld een verantwoord inzicht in het niveau van het onderzoek en op termijn in de ontwikkeling hierin.

8. SAMENVATTING EN CONCLUSIE

In dit artikel is in een aantal stappen beschreven hoe het proces van kwaliteitsverbetering en professionalisering bij de onderzoeksbureaus is verlopen en welke discussies er momenteel spelen. Wij geven een schematisch samenvattend overzicht.

jaar	organisatie	onderwerp
1978 – 1988	WBO	minimumeisen controle veldwerk minimumeisen steekproef- en non-respons verantwoording
1988 – 1991	VMO	VMO-kwaliteitseisen (= randvoorwaarden)
1991 – 1996	VMO	telkens update van kwaliteitseisen
1993 – 1996	VMO	opstart ISO-9001
1995 – 2000	VMO	ISO-certificering van bureaus
1996 – 2000	VMO	ISO + VMO kwaliteitsregels = ISO plus
1998	VMO	Total Quality Management
2000 – 2001	MOA	naast formele eisen ook inhoudelijke beoordeling

De bescheiden agenda die de vergaderingen van de Werkgroep Bureau Overleg in de beginjaren kende staat in schril contrast met de veelheid van onderwerpen die in de MarktOnderzoekAssociatie aan de orde zijn. Steeds meer wordt er aandacht geschonken aan de professionaliteit van het vak. De toename in omvang van het onderzoek, de internationalisering, de wet- en regelgeving, de non-responseproblematiek, een toenemende concurrentie, de introductie van nieuwe interviewtechnieken en vooral de eisen die van opdrachtgeverszijde worden gesteld zijn dwingende factoren om een keuze te maken voor de kwaliteit van het vak. Een belangrijke stap daarvoor wordt nu door de Research Keurmerkgroep gezet. Daarmee staat deze organisatie overigens niet alleen. Ook de Vereniging voor Beleidsonderzoek (VBO) werkt op een vergelijkbare wijze aan het hanteren van kwaliteitsnormen.

Kortom, de marktonderzoeksbureaus zijn druk bezig met discussies en uitwerkingen met als doel de kwaliteit van het vak te verhogen, zowel in nationaal als internationaal verband. Kwaliteitszorg en professionalisering in het marktonderzoek zal de komende jaren een belangrijk onderwerp blijven.

LITERATUUR

- Berg, H. van den, M. Mehcziz, H. Houtkoop, B. Holleman, 'Marktonderzoek: opinie meten of opinie maken', VU, Amsterdam, febr. 2000
- Bronner, A.E., R. van Leeuwen, 'VMO-kwaliteitseisen', juni 1991 (update 1994, 1996)
- ESOMAR Nieuws Brief, 'Raising the Quality of Research', no 2, April 1991
- ESOMAR Nieuws Brief, 'Quality standards in market research', Nov. 1998, p.3-5
- Ferrell, O.C., S.J. Skinner, 'Ethical Behavior and Bureaucratic structure in marketing research organizations', Journal of Marketing Research, vol. 25, Febr. 1988, 103-9
- Görtemöller, G., 'Van ISO 9001 naar Total Quality Management', interne VMO notitie, jan. 1998
- Kuin, H., 'VMO-eisenlijst weerspiegelt consensus onderzoeksbureaus', Adformatie nr. 25, 20 juni 1991, p.34

Marbeau, Y., 'EFAMRO Market Research Quality Standards', EFAMRO, July 1998
Motivaction, 'Brancheonderzoek Marktonderzoekbureaus', Amsterdam, jan. 2001
MRS, 'Guide to the practice of market and survey research', 1980
Sommer, R., G. Unholzer, E. Wiegand, 'Standards zur Qualitätssicherung in der Markt- und Sozialforschung', Oktober 1999, Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute
Veldman Tj., Notitie van Commissie Toetsing en Toelating, 31 oktober 1994, gerapporteerd aan het WBO-bestuur
Veldman Tj., Uit een rapportage van de Commissie Toetsing en Toelating aan de WBO, 8 september 1997
VMO, 'Handleiding voor de opzet van een kwaliteitssysteem en een kwaliteitshandboek voor marktonderzoekbureaus', juni 1994
Westerhoven. E. van, 'Kwaliteitsnormen Marktonderzoek', WBO, nov. 1983
Westerhoven. E. van, 'Eerste notitie betreffende zelfregulering marktonderzoekburo's', WBO, mei 1984.

3. Online kwalitatief marktonderzoek: theoretische en conceptuele uitdagingen

E.A.G. GROENLAND

SAMENVATTING

Dit artikel gaat in op nieuwe werkwijzen en methodologieën voor het kwalitatief marktonderzoek, welke zijn gebaseerd op online gegevensverzameling. Ontwikkelingen die mogelijk worden door de recente technologische ontwikkelingen rondom het internet. Gesteund door de resultaten van literatuuronderzoek zal worden betoogd dat het Internet een Umfeld creëert dat voor een belangrijk deel heel eigen effecten heeft op de wijze waarop respondenten en de moderator met elkaar communiceren. Op grond hiervan worden voorstellen gedaan voor aangepaste vormen van online kwalitatieve onderzoeksmethodieken.

Ter illustratie van de kracht en de potentie van online kwalitatief marktonderzoek zullen de resultaten van een vergelijkend onderzoek tussen klassieke en nieuwe vormen worden gepresenteerd. De resultaten tonen aan dat, onder bepaalde voorwaarden, en voor bepaalde typen vraagstellingen, dedicated chatprogramma's en bulletin boards met vrucht kunnen worden ingezet bij de beantwoording van onderzoeksvragen.

De tekst wordt besloten met een vooruitblik van de uitdagingen, de mogelijkheden en de begrenzingen die de nieuwe technologie, en de virtuele communicatiemogelijkheden tussen bureau en respondent ons, naar verwachting, zullen gaan bieden.

1. INLEIDING

Het Internet wint in toenemende mate aan belang als informatiebron, als marktplaats, en als communicatiemiddel. Voor consumenten en burgers geldt dat, in aanvulling op de fysieke wereldbeleving, de virtuele wereld gaandeweg aan belevingskracht wint. In dit verband stelt Jones (1999) dat consumenten en burgers in de 21^e eeuw zowel in 'physical space' als in 'cyberspace' leven. Hierbij vallen traditionele grenzen met betrekking tot tijd, ruimte, en afstand grotendeels weg, zeker in de beleving van individuen. Dit gaat gepaard met veranderingen in het gedragsrepertoire, zowel wat betreft 'online-' als 'off-line'-gedrag. Ook individuele ervaringen, identiteit en zelfbeeld ontwikkelen zich, en wel in een richting waarbij mogelijk zelfs een zekere versmelting plaatsvindt van de beide realiteiten.

Marktonderzoek stelt zich ten doel de realiteitswerelden van burgers en consumenten empirisch te onderzoeken, en de consequenties van hun behoeften, wensen, belevingen, voorkeuren en oordelen op hun (koop)gedrag te beschrijven en te voorspellen. De

door consumenten en burgers ervaren realiteit dient als uitgangspunt te worden genomen voor het ontwikkelen van theorieën over online onderzoek (Jones, 1999). Waar de belevingswereld van deze doelgroepen inhoudelijk verandert, dienen ook de theorieën en meetinstrumenten te worden aangepast. Dit geldt a fortiori voor het kwalitatief markt- en beleidsonderzoek, waarbij individuele processen en subjectieve realiteitsbelevingen centraal staan.

Tegen de achtergrond van deze ontwikkelingen zullen we de potentie van het Internet, als medium voor het verkrijgen van empirische marktgegevens, bespreken. Hierbij staan de volgende vragen centraal:

- Welke kenmerken en eigenschappen heeft het nieuwe medium in de perceptie en de beleving van consumenten en burgers?
- Hoe verlopen communicatieprocessen binnen het Umfeld van het Internet?
- Welke –bestaande en nieuwe- kwalitatieve onderzoeksmethodieken zijn functioneel en effectief in de virtuele online omgeving?
- Hoe verhouden traditionele kwalitatieve onderzoeksmethodieken zich tot online onderzoeksmethodieken?

We baseren ons bij de beantwoording van deze vragen in eerste instantie op de wetenschappelijke literatuur¹. Daarna presenteren we enige empirische bevindingen uit het aan de leerstoel² verbonden online kwalitatief onderzoeksprogramma³.

2. HET INTERNET

Het begrip ‘Internet’ is van een inherente abstractie, aangezien het verwijst naar een virtuele wereld. Dat wil zeggen: een wereld die niet door middel van onze vijf zintuigen is waar te nemen. Anderzijds vertegenwoordigt de virtuele wereld een nieuwe omgeving voor burgers en consumenten die rijk is aan persoonlijke indrukken, en belevingen: in psychologisch opzicht kunnen we spreken van een wereld met een grote en rijke psychologische inhoud. Dit blijkt onder meer wanneer we het Internet beschrijven en typeren vanuit de belevingswereld van deze gebruikers.

2.1. Kenmerken

Bezien we de kenmerken die gebruikers noemen, dan doemt het beeld op van een geheimzinnige, spannende, nieuwe wereld, of ruimte die is te betreden. Deze wereld heeft zijn eigen domeinen, zijn eigen bewoners, en zijn eigen spelregels. Op de achtergrond is nog iets van de vrijbuitery, het alternatieve, en het anarchistische te proeven waarmee de ontstaansgeschiedenis van het nieuwe medium is geassocieerd.

¹ Ik ben Joris de Rooij zeer erkentelijk voor de assistentie die hij mij heeft geboden bij de uitvoering van het literatuuronderzoek.

² Dit betreft de door Blauw Research ingestelde leerstoel “Qualitative Business Intelligence” aan Universiteit Nyenrode.

³ Grote dank ben ik verschuldigd aan Marc Scholten, die een majeure bijdrage heeft geleverd aan de uitvoering en rapportage van het te presenteren online-onderzoeksproject.

Een deel van de fascinatie van het Internet is te verklaren doordat communicatieprocessen in het algemeen, en contacten met anderen in het bijzonder, op andere wijze verlopen, en aan andere wetten schijnen te gehoorzamen. Dit geldt uiteraard ook voor het proces van groepsvorming dat het gevolg is van communicatie. Hierbij blijkt het bij voorbeeld mogelijk te zijn een andere (tweede) identiteit aan te nemen, en een nieuwe invulling te geven aan het zelfbeeld. Opvallend is dat de fysieke en de virtuele wereld hierbij gescheiden entiteiten zijn, en blijven.

Gebruikers typeren het Internet in de eerste plaats door een aantal kenmerken te noemen (zie tabel 1).

Tabel 1. Internet naar kenmerken.

-
1. Virtueel, dus ontastbaar
 2. Geen restricties met betrekking tot afstand en tijd
 3. Het groeit en verandert
 4. Communicatiekanalen (hyperlinks), creëren, vernietigen
 5. Geen bezit op het internet: je deelt informatie met de 'community'
 6. Politiek-ideologisch: community is democratisch, vrij, geen gezagsverhoudingen, gratis
 7. Betreden van deze wereld naar eigen keuze
-

De virtuele wereld is een nieuwe wereld, anders dan de fysiek-sociale wereld. Er zijn nieuwe regels te ontdekken, nieuwe doeleinden te formuleren, en nieuw gedrag vast te stellen.

2.2. Communicatiemedium

Het Internet kent twee vormen van communicatie. In de eerste plaats betreft dit het onpersoonlijke eenrichtingverkeer. Hier wordt het medium gebruikt voor het aanbieden en opnemen van enorme hoeveelheden informatie door de gebruikers. In aanvang betrof dit voornamelijk informatie van wetenschappelijke en maatschappelijke aard. De vercommercialisering van het Internet houdt in dat er een toenemend aanbod van commerciële 'sites' (virtuele locatie op het Internet) valt waar te nemen. Een voorbeeld uit de business-to-consumer-markt moge dit illustreren. Zo beschrijven Krekt et al. (1999) pogingen om, via commerciële Internetsites, emotioneel geladen informatie over te dragen aan consumenten, en bespreken in dit verband het begrip 'belevingscommunicatie'. Ze definiëren dit als een verzameling van door merken geënceneerde, belevingsgeoriënteerde gebeurtenissen die de consument op eigen initiatief ondergaat. Een tweede vorm van communicatie betreft het interpersoonlijke tweerichtingsverkeer. Het Internet maakt het mogelijk om zowel synchroon met elkaar in verbinding te staan, als op asynchrone wijze te communiceren. In het eerste geval gebeurt dit met name door middel van zogenaamde 'chat-rooms'. In het laatste geval maken partijen gebruik van E-mail, of bulletin boards in de vorm van nieuws- of discussiegroepen. Kenmerkend aan deze vorm van communicatie zijn de mogelijkheden, en het gemak, waarmee de connectie kan worden gelegd, en weer verbroken.

De virtuele wereld is een nieuwe wereld, met nieuwe kansen en mogelijkheden om contacten te leggen, en te verbreken, en andere mensen te ontmoeten. Onderzoek zal moeten uitwijzen hoe de dynamiek van deze processen verloopt.

2.3. Duiding en beleving

In de beleving van de gebruikers biedt het Internet geheel eigen mogelijkheden om individuele relaties aan te gaan. Ook kan men er voor kiezen om deel uit te gaan maken van een grotere, sociale eenheid: de 'community'. 'Community' is hierbij op te vatten als een nieuw construct voor de deelnemers. Het begrip 'community' is abstract, en vervloeit en vervormt in de beleving en de verbeelding van de deelnemers. De waargenomen community-identiteit is gebaseerd op eigen beelden en indrukken, en voor een deel ook ronduit op de eigen fantasie. Deze beelden en indrukken ontstaan, en ontwikkelen zich tijdens bij voorbeeld chat-sessies. Het feit dat de deelnemers anoniem kunnen blijven, als ze dat wensen, versterkt nog dit proces.

Het ongrijpbare karakter van het Internet, met zijn communities, wordt geïllustreerd door het veelvuldige gebruik van metaforen. Het gebruik van metaforen kan worden opgevat als (niet-bewuste) pogingen van Internet-gebruikers om vat te krijgen op deze nieuwe wereld.

Tabel 2. Metaforen voor het Internet.

Ruimtelijke metaforen	Markt-metaforen
1) global village	1) global market place
2) digitale stad	2) elektronisch marktplein
3) Spaceship earth	3) world economy
4) cyberspace	

Communicatie metaforen
1) elektronische snelweg

Een analyse van de gebruikte metaforen laat zien dat het Internet wordt beleefd als elektronische landsgrens, als markt, als sociale omgeving, en als luik naar de gemeenschap. Het kenmerk van metaforen is dat ze alle elementen van vergelijkbaarheid bezitten, maar geen van alle de belevingsdimensies van de gebruikers afdekken.

Onderzoek zal moeten uitwijzen tot welke belevingsdimensies het Internet aanleiding geeft, bij de diverse onderscheiden typen gebruik, en gebruikers.

2.4. Virtuele versus materiële aspecten

Het vorenstaande levert een algemeen beeld op waarbij gebruikers het Internet ervaren als een nieuwe wereld, die 'anders' is in termen van gebruik, beleving, regels en mogelijkheden. Het zet hen er toe aan om te experimenteren met hun identiteit, hun zelfbeeld, door het zoeken en onderhouden van contacten, om daarmee een eindwaarde als geestelijke groei (Maslow, 1954) te bevredigen. In die zin kan het Internet worden gezien als een springplank naar een meer vergeestelijkt en spiritueel bestaan. Benedikt (1991) spreekt in dit verband van 'etherealization' van onze wereld, een proces waarbij het menselijk bestaan zich verheft naar hogere energievormen. In die zin kan worden gesteld dat het Internet mogelijkheden biedt tot verbreding en verdieping van het menselijk functioneren.

De toekomstvisie van Sheth & Sisodia (1997) blijft dicht bij materiële, aardse zaken,

en richt zich op de veranderingen in consumentengedrag onder invloed van het Internet. De auteurs schetsen op basis van een door hen uitgevoerde meta-analyse een toekomstbeeld waarbij consumenten een ander gebruik gaan maken van hun vier schaarse hulpbronnen: geld, tijd, inspanning, en ruimte.

Tabel 3. Hulpbronnen en Internet.

-
- 1) Geld: consumenten verwachten meer waar voor minder geld
 - 2) Tijd: consumenten zijn bereid om meer te betalen teneinde tijd te winnen
 - 3) Inspanning: consumenten in een complexe wereld zoeken gemak en eenvoud
 - 4) Ruimte: grootgebruikende consumenten wensen kleine hoeveelheden product voor weinig geld door te profiteren van economy of scale-effecten
-

Al deze voorbeelden zijn gebaseerd op de centrale plaats en het gebruik van het Internet in het leven van consumenten.

De wetmatigheden in het consumentengedrag, alsmede de determinanten ervan, dienen opnieuw beschreven te worden. Hierbij dienen zowel materiële als immateriële gedrags- en omgevingsaspecten te worden betrokken. Onderzoek dient nader inzicht te verschaffen in de onderlinge verwevenheid van de virtuele en de fysiek-sociale werelden van consumenten en burgers, en de verschillende belevingsdimensies die deze werelden oproepen.

We constateren dat het consumentenleven zich deels afspeelt in de virtuele omgeving, waar consumenten hun opvattingen en zienswijzen deels ontleen aan het gebruik van, respectievelijk het omgaan met het Internet. Volgens het principe van de ecologische validiteit is het prudent om empirisch marktonderzoek en gegevensverzameling (deels) uit te voeren binnen het nieuwe Umfeld van consumenten en burgers. Dit Umfeld zal in de volgende paragraaf nader worden beschreven.

3. UMFELD

Het Umfeld kan worden opgevat als een omgeving, een context, waarbinnen in het algemeen gedrag plaatsvindt. Bij het uitvoeren van marktonderzoek vindt de dataverzameling altijd plaats binnen een bepaald Umfeld. Traditionele assessmentomgevingen in het marktonderzoek worden gevormd door:

Tabel 4. Kenmerken van traditionele assessmentomgevingen.

Traditionele omgevingen in het marktonderzoek worden gevormd door:

- 1) de individuele persoon, met zijn binnenwereld, die een postale enquête invult
 - 2) de individuele persoon met de directe ander (dyade), waarbinnen het face-to-face-interview of het diepte-interview plaatsvindt
 - 3) de individuele persoon met de indirecte ander (dyade), waarbinnen het telefonisch interview plaatsvindt
 - 4) individuele persoon met de directe anderen, waarbinnen het groeps gesprek plaatsvindt
-

Door de opkomst en de verbreiding van het Internet kunnen hier aan worden toegevoegd, zie tabel 5.

Tabel 5. Kenmerken van virtuele assessmentomgevingen.

Virtuele omgevingen in het marktonderzoek worden gevormd door:

- 1) de individuele persoon, met zijn binnenwereld, die een E-mail-enquête invult
 - 2) de individuele persoon met de indirecte, synchrone ander, waarbinnen het individueel diepte-interview, of de interpersoonlijke chat plaatsvindt
 - 3) de individuele persoon met de indirecte, synchrone anderen, waarbinnen de discussiegroep actief is, of groeps-chat plaatsvindt
 - 4) de individuele persoon met de indirecte, asynchrone anderen, waarbinnen activiteiten rondom het bulletin board plaatsvinden
-

De definitieverschillen hebben in deze indeling betrekking op de binnen- en buitenwereld van de individuele persoon, het directe of indirecte contact, en het aantal personen waarmee wordt gecommuniceerd.

Vanuit een sociaal-psychologisch perspectief bezien gaat het hier steeds om metingen in assessmentomgevingen waarvoor in het algemeen geldt dat respondenten, maar ook interviewers, een bepaalde rol spelen, en bepaalde verwachtingen hebben ten aanzien van hun rolinvulling. Voor de onderscheiden omgevingen geldt dat de aard, de inhoud en de intensiteit van het contact en de interacties tussen de partijen in het sociaal proces verschillend zijn, en dat hiermee elke dataverzamelmethode zijn eigen meetfouten, en vertekeningen introduceert. Een belangrijk verschil tussen kwalitatief en kwantitatief onderzoek is dat bij het eerste type onderzoek de onderzoeker een eigen, processturende invloed uitoefent. Waar de kwantitatief interviewer wordt getraind om geen eigen invloed uit te oefenen tijdens de gegevensverzameling (anders dan een algemeen motiverende) staat de kwalitatief onderzoeker voor de uitdaging om, naast elicitatietechnieken, onder meer zijn of haar eigen persoonlijkheid in te zetten om diepere inzichten te verkrijgen op basis van de reacties van de respondent.

Bezien we de literatuur rondom het Internet als medium voor het uitvoeren van onderzoek, dan kan het Internet-Umfeld worden getypeerd door middel van de volgende dimensies.

Tabel 6: Dimensies van het Internet-Umfeld.

-
- 1) Reductie van de interpersoonlijke, zintuiglijke waarneming.
 - 2) Veranderingen ten aanzien van de perceptie en beleving van de tijd-ruimte-dimensie.
 - 3) Veranderingen ten aanzien van de ervaren, interpersoonlijke afstand.
 - 4) Het ontstaan van 'communities'⁴.
 - 5) Nieuwe regels en symbolen ten behoeve van de interpersoonlijke communicatie.
-

⁴ Theoriegestuurde, kwalitatieve onderzoeksmethoden vallen buiten het bestek van deze verhandeling. Uiteraard lenen deze benaderingen zich bij uitstek voor de studie van het ontstaan, de groei en de verandering, en de neergang van virtuele communities. Hetzelfde geldt voor thema's als omgangsvormen, rituelen, communicatievormen, taalgebruik, sociale beloning en straf, cultuur, positieve en negatieve emoties, ideologieën, levensstijlen, community-geschiedenis, en het collectieve community-geheugen.

Het Internet is een virtuele omgeving, waarin het zintuiglijke zien en horen niet langer bestaat. De waarneming geschiedt vrijwel uitsluitend via het scherm van de computer-monitor. Zoals de telefoon een reductie van zintuiglijke waarneming impliceerde ten opzichte van het face-to-face-gesprek, zo impliceert het Internet een reductie van zintuiglijke waarneming ten opzichte van de telefoon.

Dit geldt in versterkte mate voor de zogenaamde 'off-line-interacties, waarbij de contacten niet synchroon verlopen. Als gevolg hiervan ontstaan gevoelens van anonimiteit en vervreemding. Deze hebben effecten op de wijze waarop een individu zijn of haar binnenwereld ervaart.

Hierbij verandert de psychologische afstand, als tegenhanger van de fysieke afstand. Er is sprake van virtuele, aanwezigheid, waarbij asynchrone contacten zijn te vergelijken met vriendschapscorrespondentie, of correspondentie-schaak. Deze beleving beïnvloedt de gevoelde vertrouwelijkheid, de aard van het interpersoonlijk contact, en de ervaren intimiteit. Als gevolg hiervan zal de aard en inhoud van de communicatie wezenlijk veranderen.

Processen van groepsvorming en groepsidentiteit zullen in aanvulling op die van de fysiek-sociale wereld, plaatsvinden. De beleving van een groepsgevoel, de identificatie met een Internet-community, de ervaring van verbondenheid met virtuele groepsleden vormen onderdelen van dit proces. Enerzijds zullen deze processen de wetmatigheden van de psychologie van de kleine groep volgen. Anderzijds kunnen aan dit proces unieke kenmerken worden onderscheiden, in termen van regels, gewoontes, rituelen, en gedragsmodaliteiten. In sommige gevallen zal er sprake zijn van complementariteit, in andere gevallen van substitutie.

Internetgebruikers zullen een nieuwe taal moeten leren. Deze taal, die overigens aan voortdurende veranderingen onderhevig is, heeft een gedeeltelijk andere syntax en grammatica, en kenmerkt zich door een enigszins ander woordgebruik. Het bedient zich van specifieke symbolen voor het overbrengen van emoties, abstracties en beelden.

Over een aantal van deze Umfeld-dimensies zijn empirische onderzoeksresultaten bekend. Deze zullen nu worden besproken.

3.1. Waarden, normen, en role playing

Een community is een wereld op zich, met eigen regels, gebruiken en rituelen. Het virtuele karakter noopt, zoals reeds besproken, tot het gebruik van metaforen. Hiermee wordt het voor de deelnemers gemakkelijker om een mentaal beeld op te bouwen van de situatie. Een bekende metafoor is het beeld van kamers, en ruimten, elk met hun eigen functie.

Het Internet-Umfeld biedt relatief veel vrijheidsgraden voor het tonen van extreme gedragsmodaliteiten. De nieuwe wereld van het Internet gaat voor veel gebruikers gepaard met het mentaal loslaten van de regels, conventies, en omgangsvormen van de oude wereld. Ook de anonimiteit waarmee men deze nieuwe wereld kan betreden, en verlaten, is hier debet aan. Het bestraffen van ongewenste of niet gewaardeerde gedragsuitingen is aan een vluchtig, en momentaan, virtueel krachtenveld onderhevig. De gebruiker kan namelijk met één druk op de knop de virtuele wereld, met haar virtuele bewoners, en de virtuele gedragscorrectie laten verdwijnen. De fysiek-sociale

wereld kent in dit opzicht een hogere mate van permanentie. Dit heeft een remmende werking op het voorkomen van extreme gedragingen.

Het Umfeld van het Internet kent een aantal dimensies welke wezenlijk verschillen van die van de fysiek-sociale wereld. Deze verschillen hebben onder meer betrekking op de interpersoonlijke oriëntatie, regels, gebruiken, conventies, houdingen, verwachtingen, en gedragingen van hen die deze nieuwe wereld betreden en bewonen. Onderzoek zal moeten uitwijzen hoe de processen van interpersoonlijke betekenisgeving in dit Umfeld verlopen, en hoe deze dienen te worden geïnterpreteerd.

4. INTERNET-COMMUNICATIE

Een Cybercommunity is zowel een entiteit als een proces. Het combineert kennis, culturele waarden en normen van de 'community' met noties over cyberspace. Het is een arena waarin passies opvlammen, en worden uitgeleefd, problemen worden opgeworpen en opgelost, sociale contacten ontstaan, en worden onderhouden, macht en controle worden uitgeoefend, de grote, existentiële problemen van het mens-zijn worden besproken, ongelooflijke trivialiteiten met ernst en geestdrift aan de orde komen, kongsi's en coalities worden gesmeed en verraden, kortom waarin wordt geleefd. (vrij naar Fernback, 1999).

De verschillen tussen face-to-face-communicatie en virtuele communicatie ontstaan door het feit dat bij de virtuele communicatie gebruik wordt gemaakt van een computer-interface. Dit Umfeld-kenmerk wordt door Döring (1997) 'computer mediated communication' (CMC) genoemd. Döring analyseert deze zogenaamde CMC-omgeving, en detecteert een negental Umfeld-invloeden die het karakter en de beleving van Internet-communicatie beïnvloeden en veranderen. Deze invloeden worden gepresenteerd als benaderingswijzen, en kunnen worden opgevat als alternatieve opvattingen over de effecten van het Internet-Umfeld. Ze kunnen als volgt worden beschreven:

Tabel 7. Alternatieve opvattingen over het Internet-Umfeld.

-
1. CMC leidt tot een reductie van communicatiekarakteristieken in het communicatiekanaal.
 2. CMC leidt tot het uitfilteren van sociale cues.
 3. CMC leidt tot een beredeneerde en geoptimaliseerde mediumkeuze en mediumvoorkeur.
 4. CMC leidt tot een normatieve mediumkeuze.
 5. CMC leidt tot een interactieve mediumkeuze.
 6. CMC leidt tot effectieve sociale informatieverwerking.
 7. CMC leidt tot de mogelijkheid om de eigen persoon te simuleren, of als een ander personage te presenteren.
 8. CMC leidt tot de mogelijkheid om te fantaseren, en tot de beleving een andere persoon te zijn.
 9. CMC is een combinatie, of een vermenging van mondelinge en schriftelijke communicatiekenmerken en -karakteristieken.
-

We lichten deze Umfeld-invloeden successievelijk toe.

1. CMC leidt tot een reductie van communicatiekarakteristieken in het communicatiekanaal.

De consequentie is een verarming en een beperking van de communicatie. De beperkingen in de interpersoonlijke contacten betreffen de zingeving en de werkelijkheidsbeleving, de beleving en de uitwisseling van emoties. Deze leiden per ultimo tot een verzakelijking en een ontmenselijking in de contacten.

2. CMC leidt tot het uitfilteren van sociale cues.

De ervaren anonimiteit in het communicatieproces leidt tot gevoelens van ontremming, en daarmee tot een toename van zowel pro- als anti-sociaal gedrag.

3. CMC leidt tot een beredeneerde en geoptimaliseerde mediumkeuze en mediumvoorkeur.

Het best passende medium bij een bepaald type boodschap wordt gezocht, en gevonden. Deze optimale combinatie leidt tot een verrijking van de communicatie. De kosten-batenafweging door zender maakt de communicatie efficiënt.

4. CMC leidt tot een normatieve mediumkeuze.

De mediumkeuze wordt bepaald door sociale normen en opvattingen in de omgeving, en volgt een model voor sociale beïnvloeding. Dit leidt vaak tot irrationeel of disfunctioneel mediumgebruik.

5. CMC leidt tot een interactieve mediumkeuze.

Door leerprocessen laten gebruikers hun mediumkeuze in toenemende mate leiden door verwachtingen, gewoontes, en eerdere mediakeuzes die goed of plezierig uitpakten.

6. CMC leidt tot effectieve sociale informatieverwerking.

CMC is net zo communicatierijk als face-to-face-communicatie, dit omdat alle non-verbale cues zijn te verbaliseren, ook al kost dit meer tijd. Te denken valt aan het gebruik van emoticons, smileys, frowny, en speciale woorden die emoties uitdrukken. Dit laatste gebeurt door het letterlijk weergeven van een woord, het gebruik van zeer lange woorden, en het schrijven van woorden in hoofdletters. De veronderstelling is dat mensen in staat zijn gevoelens over te brengen, relaties tot stand te brengen, en sociale vaardigheden in het Internet-Umfeld te leren. Storingen ontstaan wanneer deelnemers ongeïfend zijn, de codes niet kennen, elkaar niet kennen, of onder tijdsdruk staan. Alle tekortkomingen zijn echter compenseerbaar. De communicatie wordt in het algemeen vergemakkelijkt wanneer een zender een foto van zichzelf, van familie, of van vrienden laat zien, iets van zijn achtergrond vertelt, iets persoonlijks communiceert, of zich kwetsbaar opstelt.

7. CMC leidt tot de mogelijkheid om de eigen persoon te simuleren, of als een ander personage te presenteren.

Er zijn vrijheidsgraden om jezelf in de communicatie anders te presenteren aan de ander. Het is mogelijk om bepaalde, plezierige of ideale eigenschappen naar voren te halen, of jezelf op frauduleuze wijze als waarlijk anders te presenteren. Dit is ook een manier om te experimenteren met eigen, mogelijke identiteiten en rollen, en kan leiden tot zelf-reflectie en bezinning.

8. CMC leidt tot de mogelijkheid om te fantaseren, en tot de beleving een andere persoon te zijn.

Door het ontbreken van zintuiglijke waarneming (zien, horen, voelen, ruiken) wordt de andere persoon ‘ingevuld’. Er kan sprake zijn van evocatieve processen, projecties van het zelf, illusies, en dromen. Hierdoor is het mogelijk om, tijdelijk, en binnen het Internet-Umfeld, boven de eigen beperktheid uit te stijgen.

9. CMC is een combinatie, of een vermenging van mondelinge en schriftelijke communicatiekenmerken en –karakteristieken.

Deze omvatten het onmiddellijke, het synchrone (chat), en het tijd- en plaatsgebonden van bij voorbeeld het concipiëren van de boodschappen. De kenmerken worden gecombineerd met het vastleggen van boodschappen, met de mogelijkheid er op te reflecteren, en het manipuleren en transporteren van deze boodschappen door middel van files en hyperlinks. De combinatie van karakteristieken heeft gevolgen voor de stijl en het ritme waarmee de communicatie plaatsvindt, alsmede de mediumkeuze.

De bovenstaande invloeden van het Internet-Umfeld kunnen worden opgevat als hypothesen. Op dit moment staat niet vast welke van de beschreven effecten zich ook daadwerkelijk manifesteren. Ook is ongewis voor wie ze gelden, en onder welke voorwaarden. Vergelijkend onderzoek dient te leiden tot een specificatie van deze factoren, alsmede de voorwaarden waaronder ze zich voordoen. Nu al is echter duidelijk dat commercieel kwalitatief marktonderzoek rekening dient te houden met het voorkomen van Umfeld-factoren, en de vertekende effecten die ze kunnen hebben op de verzamelde gegevens. De uitdaging zal zijn deze factoren te onderkennen, en aangepaste werkwijzen te ontwikkelen die efficiënt, effectief en uitvoerbaar zijn.

5. ONLINE KWALITATIEVE ONDERZOEKSMETHODEN

5.1. Inleiding

Online kwalitatief onderzoek vindt momenteel met name in de Verenigde Staten plaats. De meest voorkomende vorm is de online groepsdiscussie; onderzoeksfora worden minder vaak ingezet. Een belangrijk argument om het instrument van de Online focusgroepen in te zetten is dat deze ‘kosten-effectief’ kunnen worden ingezet: ze leveren veel informatie voor verhoudingsgewijs weinig geld. Een additioneel voordeel betreft het feit dat dit instrument het mogelijk maakt om zogenaamde ‘low incidence groups’, of ‘specialized audiences’ te bereiken. Opvallend is hierbij de hoge mate van gestructureerdheid: een groot aantal vragen passeert de revue tijdens een sessie, waarbij verhoudingsgewijs weinig gelegenheid wordt genomen om tot diepgang te komen. Een relatief groot aantal respondenten werkt zich met grote snelheid door de vragen heen. De vragen hebben een semi-kwantitatief karakter. Efficiency lijkt het te winnen van het tijdrovende verbeteren van de kwaliteit van de antwoorden. Gesteld kan worden dat in wezen de klassieke methode van de groepsdiscussie is getransponeerd naar de online omgeving, zonder expliciet rekening te houden met de mogelijkheden en de beperkingen van het Internet-Umfeld.

Om eerder besproken redenen ware het beter hier uitdrukkelijk aandacht aan te beste-

den. Dit doen we door een nieuwe indeling van online onderzoeksmethoden te presenteren, met aandacht voor de taken en functies van de moderator bij elk van deze onderzoeksmethoden.

De eerder besproken studies laten zien dat de oervormen van communicatie via het Internet als zodanig niet optimaal geschikt zijn om ze te implementeren in het commercieel onderzoek. Enerzijds lijken deze zogenaamde ‘natuurlijke’ communicatievormen bij uitstek geschikt voor participerende observatie. Anderzijds ontberen ze met name de taakgerichte sturing, en de sociaal-emotionele motivering van de online moderator, zoals de discussieleider of de interviewer in de ‘nonline’- (ofwel off-line) onderzoeksmethoden. Ook de vrijblijvendheid van deelname, en het ontbreken van de verplichting voor deelnemers, een prestatie te leveren bij participatie in de ‘natuurlijke’ chat (N-chat), of het ‘natuurlijke’ bulletin board (NBB) doen afbreuk aan de effectiviteit van de sessies. Daarom dienen voor het commercieel kwalitatief marktonderzoek aangepaste onderzoeksmethodieken te worden ontwikkeld.

Online onderzoeksmethoden kunnen worden ingedeeld naar:

- One-to-one versus One-to-many met betrekking tot het aantal onderzoeksdeelnemers
- Synchroniciteit versus asynchroniciteit met betrekking tot het proces van dataverzameling

Toepassing van de bovengenoemde indelingscriteria, en de introductie van de moderator leiden tot een typologie waarin de Online Single, de Online Letter, de Online Group, en het Online Forum kunnen worden onderscheiden. Deze vervangen respectievelijk het kwalitatieve diepte-interview, bepaalde varianten van de kwalitatieve Delphi-methode, de focusgroep, en het bulletin board. Tabel 8 toont deze indeling in de vorm van een vierveldenmatrix.

Tabel 8. Een prototypische indeling van online onderzoeksmethoden.

	Aantal Onderzoeksdeelnemers	
Proces van dataverzameling	One-to-one	One-to-many
Synchroon	Online Single	Online Group
Asynchroon	Online Letter	Online Forum

Bij elk van deze basistypen dient de rol van de interviewer/moderator nader te worden bepaald, en wel op basis van vergelijkende, empirische studies. De eerste fase van het onderzoeksprogramma bestaat dan ook uit een inventarisatie en exploratie van de moderator-invloeden bij de onderscheiden online onderzoeksmethoden. Voordat we de onderzoeksresultaten van ons eigen onderzoeksprogramma presenteren bespreken we een aantal gerapporteerde bevindingen over enige van deze online onderzoeksmethoden. Hierbij maken we gebruik van de nieuwe, door ons ingevoerde labels. De te bespreken onderzoeksuitkomsten zijn uitgesproken schaars, en fragmentarisch. Ze vormen echter een eerste uitgangspunt voor het opzetten en uitvoeren van exploratief vervolgonderzoek, en vormen de opmaat voor ons onderzoeksprogramma.

Achtereenvolgens komen de Online Group, de Online Single, en het Online Forum

aan bod. Voor zover ons bekend zijn er geen empirische onderzoeken gerapporteerd met betrekking tot het instrument van de Online Letter.

5.2. Online Group

Mann en Stuart (2000) beschrijven de taken, de wenselijke houding en het wenselijke gedrag van de moderator bij het ondervragen van online groepen.

Een eerste taak verwijst naar het proces van ‘rapport’ bouwen tussen de deelnemers. Hierbij hangt de mate van de op te roepen interactie tussen de deelnemers onderling en tussen deelnemers en de moderator af van het onderwerp, en ook de stijl van de moderator. De beginfase is, net als bij een groepsdiscussie, cruciaal. In deze fase trekt de moderator het initiatief naar zich toe. Zij/hij initieert het ‘voorstelrondje’, en geeft instructies wat betreft de regels volgens welke de discussie zal worden gevoerd. Ook zorgt de moderator ervoor dat er een plezierige, niet-controversiële sfeer ontstaat. De moderator stelt duidelijke regels, en stelt in, respectievelijk handhaaft duidelijke vormen van gedragscontrole over de respondenten.

De basisdimensies van het proces zijn, naar de mening van de beide onderzoekers, vergelijkbaar met die in een face-to-face-situatie. Mann en Stuart (2000) noemen in dit verband de volgende aspecten:

Tabel 9. Aspecten van communicatie.

-
- 1) Vertrouwen opbouwen
 - 2) open en eerlijk zijn
 - 3) je kwetsbaar opstellen
 - 4) Interactieve vaardigheden in de vorm van luisteren, onmiddellijk reageren, doorvragen, interesse aangeven
 - 5) Geruststellen, aangeven dat ‘het goed gaat’
 - 6) sensitiviteit tonen door middel van woorden
 - 7) zonodig anticiperen op reacties van de deelnemers
 - 8) contact houden met de deelnemers, dat wil zeggen: laten zien dat je de respondent volgt in zijn reacties.
-

In deze opvatting volgt het proces van ‘rapport bouwen’ dezelfde principes als die bij de groepsdiscussie. De verschillen in het Umfeld nopen tot aanpassingen die niet-structureel van aard zijn.

De communicatie binnen Online Groups kent een aantal wetmatigheden, regels en karakteristieken. Allereerst geldt dat hier persoonlijke dominantie kan worden bereikt door degene die het snelste typt. Van belang voor de aard van de communicatie zijn ook het aantal deelnemers, en dynamiek tussen de deelnemers. Voorts gelden ook hier het voorkomen van ‘pair friendships’, het optreden van groepconflicten die zich online manifesteren. Tenslotte wordt de aard en de effectiviteit van de communicatie bepaald door de mate van openheid van de deelnemers. Al deze zaken bepalen in zekere mate het succes waarmee ‘rapport’ wordt gecreëerd.

Een tweede taak betreft het structureren van het gesprek. Aangezien de start van het gesprek moeilijk is, gegeven het gebrek aan directe controlemogelijkheden, kiezen

sommige moderators er voor om meteen het inhoudelijk gesprek te beginnen. Hierdoor worden de deelnemers gedwongen, of tenminste aangespoord, om een ordentelijke uitwisseling van meningen en reacties te beginnen.

De gewenste mate van structuur bepaalt of alle vragen in een keer worden aangeboden, of dat vragen sequentieel, met telkens gelegenheid tot responderen, worden gepresenteerd. Aangezien de groepsdiscussie met name gericht op vele onderlinge interacties tussen groepsleden lijkt het prudent om een meer gestructureerd gespreksformat te kiezen, waarbij alle vragen aan het begin van het gesprek tegelijk worden aangeboden. Het diepte-interview biedt meer mogelijkheden tot beheersing van de 'flow of communication', en gedijt daarom wellicht meer bij een sequentiële aanbieding.

Het gebruik van het probing-mechanisme, dat wil zeggen: het doorvragen na een initieel antwoord van een respondent, kent een aantal inherente beperkingen. Met name de overgangen in het vraag-en-antwoordspel tussen moderator en respondent zijn moeilijker te monitoren; een moderator kan wachten op een antwoord van een respondent dat nooit zal komen.

Ook het vraagformat in het online Umfeld verdient aandacht. Hierbij dienen de volgende dichotomieën centraal te staan:

- Directe versus indirecte vragen
- Brede of smalle vragen
- Een groter of een kleiner aantal vragen

In een bredere formulering kunnen we spreken van vraagmanagement, of zelfs communicatiemanagement. Hierbij gaat het om de wijze waarop de controle over het gesprek wordt verworven en behouden, de wijze waarop richting wordt gegeven aan de discussie, en de wijze waarop en de volgorde waarin onderwerpen worden aangesneden en afgerond. Een centraal aandachtspunt hierbij is het vermijden van interviewer bias, zonder de controle over de communicatie te verliezen.

Gewenst is het ontwikkelen van een (klein) aantal kernstrategieën, of vraagmanagementstijlen, die een optimale beheersing van het gesprek mogelijk maken, en voldoende sturingsmogelijkheden bieden om gespreksabberraties het hoofd te bieden.

Empirisch onderzoek zal moeten uitwijzen welke gespreksmanagement-strategieën aan deze eisen voldoen, en op welke wijze aan deze strategieën inhoud wordt gegeven.

5.3. Online Single

Het kwalitatieve diepte-interview wordt beschreven in de context van MUD's (multi-user-dungeons /-domains), zie Utz, (1999). Een kenmerk is hier dat transcripties van de dialogen onmiddellijk voorhanden zijn, en het, zonder omwegen, mogelijk is om inhoudsanalyse op de transcripties toe te passen, eventueel met behulp van de computer. Op basis van de uitgevoerde inhoudsanalyse is het vervolgens zeer goed mogelijk om typologieën van consumenten te construeren.

De Online Single lijkt zeer geschikt om individuele emoties en gevoelens en diepere, achterliggende processen te peilen. Dit zou mogelijk zijn door onder meer de beleving van anonimiteit van respondent. Emoties die de respondent ervaart tijdens het commu-

nicatieproces in de MUD-context kunnen tekstueel worden overgebracht door het intypen van mnemonics. Dit zijn korte, voorgedefinieerde woorden en aanduidingen. De software voorziet dan in een lijst van mnemonics met, voor elke mnemonic, een uitgeschreven zin die de desbetreffende emotie uitdrukt. Deze lijst kan tijdens de sessie op het scherm van de personal computer van de ondervraagde persoon worden geplaatst door het gebruik van zogenaamde 'cookies'. Als de ondervraagde persoon een popup-scherm activeert verschijnen de mnemonics. Na bijvoorbeeld het tonen van een product klikt de deelnemer aan de Online Single de bij haar of hem opgeroepen emoties aan. Deze procedure kan worden gezien als een variant van de Photosort, waarbij de respondent foto's selecteert die passen bij een stimulus.

Goede ervaringen worden gerapporteerd in de context van concept-beoordelingen, packaging, concept-advertising, en het bespreken van beta-testervaringen met nieuwe producten (Harris, 1996). Onderzoek moet ook hier uitwijzen of, en onder welke voorwaarden, deze werkwijze effectief en uitvoerbaar is.

5.4. Online Forum

Het bulletin board vormt een nieuw, fascinerend communicatie-instrument voor het uitvoeren van online kwalitatief markt- en beleidsonderzoek. Een bulletin board vormt, voor een (meestal) kleine groep deelnemers, een medium om met elkaar op asynchrone wijze te communiceren over een of meer onderwerpen. De oorspronkelijke benaming, het (natuurlijke) bulletin board, verwijst naar een virtueel mededelingenbord waarop deelnemers hun reactie kunnen 'prikken'. De deelnemers hebben hierbij de vrijheid om zelf het aantal inlogmomenten te kiezen, alsmede het tijdstip van inloggen. De psychologie van deze groep vertoont een zekere verwantschap met die van de op synchrone wijze communicerende groep, die de chat als communicatie-instrument gebruikt. Dit wordt veroorzaakt door het feit dat het Umfeld tot op zekere hoogte gelijk is. Zo is in beide gevallen sprake van vormen van anonimiteit tussen de deelnemers, veroorzaakt door de afwezigheid van directe, fysieke contacten. Bijgevolg zijn er meer vrijheidsgraden om de eigen identiteit en de eigen persoonlijkheid anders 'in te vullen' en te kleuren, kortom een tweede persoonlijkheid te ontwikkelen welke relatief onafhankelijk is van de sociaal-fysieke persoonlijkheid. Deze Internet- of online-persoonlijkheid zal, naar verwachting, slechts gradueel verschillen binnen de Umfelden van de chat en het bulletin board.

De moderator ziet zich voor een opgave gesteld die vergelijkbaar is met die bij de chat. Zij/hij poogt de individuele leden tot een groep te smeden, teneinde de voordelen te plukken van de groepsdynamiek. Dit wordt tot stand gebracht door gemeenschappelijke taken en doelen te stellen, de gelegenheid te bieden elkaar enigszins te leren kennen, een gevoel van verbondenheid op te roepen, en een proces van wederzijdse validatie te initiëren. De groepsdynamiek leidt tot groepsidentiteit, groepscohesie, en conformiteit in termen van taken en doelen, en taakopvattingen van de groep. De deelnemers bouwen aan hun sociale identiteit, wat een motiverende werking heeft.

In een natuurlijk bulletin board (NBB) kunnen leiders en volgers worden onderscheiden, en een harde kern van deelnemers versus een losse periferie. Dit bulletin board heeft een eigen ongestuurde, en ongerichte dynamiek.

Een Online Forum daarentegen is doelgericht, de moderator is er de leider, en de deelnemers volgers. Binnen de groep deelnemers ontstaan, tijdens het verloop van de communicatie, ook verschillen wat betreft status, gepercipieerde positie, en reactieve en proactieve opstelling.

In een bulletin board in natuurlijke setting / Umfeld kan er sprake zijn van macht en privileges van sommige leden. Zij bepalen bij voorbeeld dat bepaalde postings worden verwijderd, en stellen reglementen op.

Bij Online Fora nemen macht en verantwoordelijkheid de hoedanigheid aan van sturing en motivatie. Er is immers een expliciet doel gesteld, en de moderator heeft zowel een taakgerichte als een sociaal-emotionele taak. De leden zijn voor een bepaald onderzoek en een bepaald onderwerp geselecteerd en uitgenodigd. Hierbij wordt hen impliciet een interessante, geanimeerde en boeiende meningsuitwisseling in het vooruitzicht gesteld. Tegenover de vrije deelname aan het NBB staat de gratificerende deelname aan het Online Forum, waarbij de moderator de deelnemers sociale steun verleent en psychologisch beloont en motiveert. Deelname aan het Online Forum leidt tevens tot een financiële beloning na afloop van de sessie.

Harris (1996) merkt op dat, ten opzichte van 'in-person interviewing' (het traditionele diepte-interview), het Online Forum tot minder fouten leidt. Dit wordt veroorzaakt door het feit dat er in het laatste geval geen vertaalslag naar een elektronisch bestand behoeft te worden gemaakt. Ook ontstaat minder interviewer bias. Hier staat, naar de mening van de auteur, echter tegenover dat de moderator minder aspecten van de communicatie zal waarnemen. Bovendien blijft onduidelijk wat nu precies de invloed van de computer als medium is op het communicatiegedrag, de communicatiepatronen, de vorm en de betekenis van boodschappen. Tenslotte is het denkbaar dat bepaalde typen personen, met bepaalde karaktereigenschappen, bepaalde persoonlijke behoeften, of een bepaalde communicatiestijl, worden aangetrokken door het medium.

Onderzoek moet uitwijzen of de besproken effecten zich daadwerkelijk voordoen in de context van het Online Forum. Indien dit het geval is dient te worden vastgesteld waar dit leidt tot vertekeningen in de verzamelde gegevens, hoe deze worden veroorzaakt, en hoe ze kunnen worden geredieerd.

6. ONLINE KWALITATIEVE ONDERZOEKSMETHODEN: EEN VERGELIJKENDE STUDIE

De literatuurverkenning toont aan dat, wat betreft empirische kennis en inzichten, nog relatief weinig bekend is van de effecten van het Internet-Umfeld waarbinnen respondenten hun informatie prijsgeven. Duidelijk is wel geworden dat de figuur van de moderator van centraal belang is bij het uitvoeren van een Online Group en een Online Forum. Potentiële vertekeningen in de aard en inhoud van de verzamelde gegevens hangen samen met de aard en inhoud van de moderator-deelnemers-interactie. Meer in het bijzonder hangen ze onder meer af van de wijze van presentatie naar de deelnemers toe, de wijze waarop de beeldvorming van de moderator verloopt bij de deelnemers, en de wijze waarop de moderator structureert, stuurt, corrigeert en de deelnemers motiveert.

Het onderzoeksprogramma bouwt voort op deze uitgangspunten. Het stelt zich ten doel om het Umfeld en de dynamiek rondom online kwalitatief onderzoek te inventariseren en te exploreren, teneinde te komen tot de ontwikkeling en optimalisering van online kwalitatieve onderzoeksmethodieken. Deze methoden dienen ten volle gebruik te maken van de krachtige nieuwe onderzoeksmogelijkheden die het Internet biedt, met inachtneming van de specifieke kenmerken en eigenaardigheden van het nieuwe Umfeld. Een toetssteen bij de ontwikkeling dient te zijn de afwezigheid van Internetgebonden, vertekenende invloeden. Met name is ook van belang het streven naar een hoge mate van ecologische validiteit bij de ontwikkeling van deze methoden en methodieken.

Het te rapporteren onderzoek⁵ richtte zich op de Online Group en het Online Forum, en nam hierbij als vergelijkingsbasis de klassieke focusgroep. Zowel het Online Forum als de Online Group werden twee maal uitgevoerd.

De belangrijkste doelstelling van het onderzoek was het verkrijgen van kennis en inzicht ten aanzien van de aard en de effectiviteit van het type informatie dat verkregen wordt bij deze onderscheiden online-onderzoeksvormen, en de functionele invloed van de moderator in dit proces. Naast deze inhoudelijke doelstelling was het onderzoek er op gericht om procedures bij de gegevensverzameling te verbeteren, casu quo te optimaliseren.

Het thema van de studies was “online financiële dienstverlening”. De respondenten voor het onderzoeksprogramma voldeden aan de volgende selectievoorwaarden:

Tabel 10. Selectievoorwaarden.

-
1. Beslissers zijn op het gebied van financiële dienstverlening in het huishouden;
 2. Beschikken over adequate kennis ten aanzien van “financiële dienstverlening”;
 3. Thuis, op het werk, of in de onderwijsinstelling beschikken over een Internetaansluiting;
 4. Adequate ervaring hebben met, en kennis hebben van het Internet;
 5. Wekelijks minimaal 3 keer gebruik maken van het Internet.
-

Ter bevordering van de vergelijkbaarheid werd tenslotte nog gecontroleerd voor leeftijd en geslacht.

De procedures werden, binnen elk van de meetomgevingen, zoveel mogelijk gelijk gehouden. Dit gold met name voor het aantal thema's, de volgorde van het aanbieden van de thema's, het geven van instructies, het gebruik van stellingen, de wijze van interacteren met de respondenten, en het geven van feedback. De Online fora hadden een looptijd van 7 dagen, en er waren 10 deelnemers. De Online Groups hadden een duur van 75 minuten, met 5 deelnemers.

In de rest van dit hoofdstuk presenteren een selectie van de uitkomsten, waarbij we aansluiten bij de door ons gevoerde discussie. De indeling is als volgt.

⁵ Een volledige verantwoording van de onderzoeksopzet, en een inhoudelijke onderbouwing van de interpretatie van de onderzoeksuitkomsten zal geschieden door middel van separate publicaties.

In de eerste plaats komen onderzoeksuitkomsten met betrekking tot de **inhoudelijke informatie** aan de orde. Deze splitsen we uit naar het type informatie, respectievelijk de aard van de informatie.

Vervolgens bespreken we kort enige onderzoeksuitkomsten ten aanzien van **procedure aspecten**.

Daarna voeren we **vergelijkingen** uit tussen de klassieke focusgroep en de online onderzoeksmethoden. Hierbij wordt de rol van de moderator nader belicht, de wijze waarop groepsprocessen zich voltrekken toegelicht, en het gebruik van elicitatietechnieken besproken.

Vervolgens bespreken we enige specifieke onderzoeksbevindingen in het licht van eerder gerapporteerde onderzoeksuitkomsten die hierop betrekking hebben.

We besluiten dit hoofdstuk met het formuleren van aanbevelingen.

6.1. Inhoudelijke informatie

De door de respondenten gegenereerde informatie werd inhoudelijk op twee manieren geanalyseerd: zowel naar type als naar aard.

6.1.1. Type informatie

De respondenten verstrekten in alle gevallen informatie over het thema ‘online financiële dienstverlening’. Een inhoudsanalyse van de aard van de door de respondenten verstrekte informatie-elementen leverde de volgende indeling op:

Tabel 11. Indeling naar type informatie.

-
- a. kennis of ervaring
 - b. evaluatie of oordeel
 - c. attitude of houding
 - d. gedrag of gedragsintentie
 - e. behoefte
-

Deze typen informatie werden vervolgens door 7 onafhankelijke beoordelaars ‘gescoord’ voor de onderscheiden meetomgevingen. Een analyse van het type informatie en het type meetomgeving levert een aantal belangrijke inzichten op:

Tabel 12. Inzichten uit analyse.

-
- 1) Bij beide synchrone onderzoeksmethoden zijn de verdelingen van de onderscheiden typen informatie vrijwel gelijk
 - 2) Er zijn duidelijke verschillen tussen synchrone en asynchrone methoden met betrekking tot het type informatie dat wordt gegenereerd.

- A) Asynchrone meetmethoden leveren op natuurlijke wijze informatie-elementen van het type ‘evaluatie en oordeel’, en ‘attitude en houding’. Fora roepen dit als het ware op.
- B) Synchrone methoden roepen in vergelijkbare mate informatiedimensies als ‘kennis en ervaring’, en ‘gedrag en gedragsintentie’ op. Dit wordt met name veroorzaakt door het feit dat de moderator hierop doorvraagt.

- 3) Het Online Forum roept, in vergelijking met de beide synchrone meetmethoden, met name informatiedimensies als ‘attitude en houding’ op. Het relatief hoge aantal interacties tussen respondenten in het forum is hier met name debet aan.
-

6.1.2. Aard van de informatie

Een tweede inhoudelijke analyse was gericht op de vraag of het geheel van de antwoorden van de respondenten leidt tot het bereiken van de inhoudelijke onderzoeksdoelstellingen. Met andere woorden: wat is de functionele kwaliteit van de informatie, gegeven de onderscheiden meetmethoden? Om deze vraag te beantwoorden werd de informatie herschreven, en beoordeeld in de volgende categorieën:

Tabel 13. Aard van de informatie: categorieën.

Breedte

Is de reactie in vergelijking tot de doelstelling compleet?

Zijn alle onderdelen van de doelstelling aan bod gekomen, of beperkt de respondent zich tot één of een aantal aspecten van de doelstelling?

Diepgang

Is het duidelijk waarom men tot de betreffende reactie is gekomen?

Is er zicht op achterliggende factoren?

Welke zijn de beweegredenen?

Concreetheid

Is het duidelijk wat de respondent wil zeggen?

Zijn de reacties eenduidig te interpreteren?

In hoeverre raakt de reactie de kern van de zaak?

Efficiency

Welk deel van de verkregen informatie is daadwerkelijk relevant?

In hoeverre leveren opmerkingen of reacties een bijdrage aan het behalen van de doelstelling?

Hoeveel informatie is overbodig?

Op basis van de bovenstaande beschrijvingen werden de antwoorden van de respondenten in de onderscheiden meetomgevingen ingedeeld in de vier categorieën. Dit geschiedde door 7 onafhankelijke beoordelaars, allen deskundig in de praktijk van het marktonderzoek. Op basis van hun oordelen konden gemiddelde aantallen ‘scores’ worden bepaald voor de vier informatie-aspecten en de meetomgevingen. Op basis hiervan worden de volgende bevindingen gerapporteerd.

Tabel 14. Aard van de informatie: resultaten.

Breedte

Het Online Forum leidt tot bredere informatie dan de beide synchrone meetmethoden.

Diepgang

Het Online Forum leidt tot informatie met meer diepgang dan de beide synchrone meetmethoden.

Concreetheid

Het Online Forum leidt tot concretere informatie dan de beide synchrone meetmethoden, met name de traditionele focusgroep.

Efficiency

Het Online Forum leidt tot efficiëntere informatie dan de beide synchrone meetmethoden.

De algemene conclusie moge luiden dat het Online Forum, ten opzichte van de Online Group, een groot aantal voordelen biedt in termen van het verzamelen van informatie die in alle opzichten functioneel is voor het bereiken van de onderzoeksdoelstelling.

6.2. Procedurele aspecten

Een onderlinge vergelijking van de beide online methoden en de klassieke focusgroep in termen van procedurele aspecten levert het volgende beeld op.

Tabel 15. Resultaten van de vergelijking tussen beide online methoden en de klassieke focusgroep.

-
- 1) In de Online Group is het, in vergelijking met de traditionele focusgroep, voor de moderator erg moeilijk om overzicht te bewaren en tot de gewenste informatie te komen.
 - 2) In een Online Group leveren de deelnemers een meer gelijkwaardig aandeel in de verkregen informatie dan in de traditionele groepsdiscussie.
 - 3) De hoeveelheid interactie tussen respondenten is in een traditionele groepsdiscussie niet per definitie groter dan in een online setting.
 - 4) Reacties in het Online Forum (asynchroon) zijn uitgebreider, en beter onderbouwd dan in beide synchrone discussies.
 - 5) In het Online Forum behoeft de moderator, in vergelijking tot beide andere methoden, relatief weinig input te leveren.
-

6.3. Traditioneel kwalitatief onderzoek versus onderzoek in een online setting:

kenmerken

Een aantal kenmerken van traditioneel kwalitatief marktonderzoek blijkt zich in een online setting op een andere wijze te manifesteren.

Zo blijken de rol en de gewenste activiteiten van de **moderator** sterk te verschillen tussen beide online onderzoeksmethoden, alsmede ten opzichte van traditioneel kwalitatief onderzoek. Bij het Online Forum heeft de moderator een relatief passieve rol terwijl bij een Online Group de rol van de moderator actief is, en in dit opzicht sterk lijkt op die in een traditionele groepsdiscussie. Door het synchrone karakter van een Online Group heeft het taakgerichte leidinggeven veel weg van dat in een traditionele setting. Het sociaal-emotioneel leidinggeven wordt daarentegen, met name door het ontbreken van non-verbale cues, als beduidend moeilijker ervaren. Door de asynchroniciteit van het Online Forum wijkt de rol van de moderator hierbij sterk af van die bij de synchrone onderzoeksmethoden. De inbreng van de moderator is minder intensief van aard, met name omdat deelnemers hun reacties spontaan zelf motiveren en toelichten, dus zonder dat de moderator hierom hoeft te vragen. De rol van de moderator lijkt hierdoor bij asynchroon onderzoek meer controlerend dan stimulerend van aard.

Het **groepsproces** dat optreedt bij de traditionele groepsdiscussie vinden we deels terug in een online setting. Anderzijds bestaan er ten aanzien van dit aspect duidelijke verschillen tussen het Online Forum en de Online Group alsmede tussen de online en traditionele onderzoeksmethoden. De eerste fase van het groepsproces wordt “Forming” genoemd (Gordon & Langmaid, 1988). In deze fase vinden processen plaats die leiden tot groepsvorming. Dit proces kan in beide online methoden worden waargenomen. In het Online Forum vindt, door het asynchrone karakter ervan, echter geen ‘voorstelrondje’ plaats, maar geven deelnemers deze informatie gedurende het proces prijs. De hierop volgende fase is de “storming”-fase (Gordon & Langmaid, 1988). In deze fase komt de onderlinge rangorde van de deelnemers in de groep tot stand, en worden de spelregels bepaald. Vergelijken we de chat-sessie met de traditionele

groepsdiscussie voor wat betreft de rol van de moderator die de spelregels uitlegt en toelicht, dan zien we opnieuw overeenkomsten.

Een belangrijk verschil tussen de traditionele en de online onderzoeksmethoden is het vrijwel volledig ontbreken van statusverschillen tussen de deelnemers. Door de grote anonimiteit en het asynchrone karakter van het Online Forum speelt gepercipieerde status met name hier een zeer beperkte rol. De overige fases lijken wel weer sterk op elkaar. Hierbij valt op dat “weinig nieuws toevoegen”, als kenmerk uit de laatste fase, bij het Online Forum bij sommige deelnemers al in een vroeg stadium waarneembaar is. Dit is grotendeels te verklaren doordat deze deelnemers, door het beginnen met een stelling en de ruime reactietijd die men daarop kreeg, al hun kruid min of meer al hadden verschoten.

Elicitatietechnieken die in traditioneel onderzoek veelvuldig worden toegepast lijken ook in een online setting functioneel te kunnen worden ingezet. Onder meer projectieve technieken, zoals sentence completion, lijken met name geschikt bij gebruik in Online Fora, omdat de deelnemers hierbij ruim de tijd hebben om zich zorgvuldig in een ander te verplaatsen, dan wel zinnen zorgvuldig aan te aanvullen. Daarnaast zal een sterker beleefde anonimiteit in de online setting ertoe bijdragen dat deelnemers gemakkelijker hun medewerking aan deze procedures zullen verlenen. Associatieprocedures lijken met name zinvol gebruikt te kunnen worden in Online Groups, omdat de deelnemers hierbij minder denktijd hebben en dus spontaner zullen reageren. Bij Online Fora, zoals in dit onderzoek werden uitgevoerd, konden deelnemer hier rustig de tijd voor nemen. Door nieuwe technologische toepassingen kan de “denktijd” die een deelnemer ter beschikking krijgt in de toekomst ook in Online Fora worden beperkt. Rankingprocedures lijken voor beide online-methoden geschikt, maar bieden met name bij Online Fora toepassingsmogelijkheden. Voordat rankingprocedures automatisch uitgevoerd kunnen worden zal echter eerst een aantal technische aanpassingen moeten worden gerealiseerd.

6.4. Desk research en onderzoeksbevindingen

Een aantal onderzoeksbevindingen sluit, qua onderwerp, aan bij specifieke hypothesen en verwachtingen uit de literatuur. Deze zullen nu worden besproken.

6.4.1. Lagere kosten

Volgens Aoki (1999), Burke (2001) en Nucifora (2000) zouden de verhoudingsgewijs lagere kosten het grote voordeel zijn bij het toepassen van online kwalitatief marktonderzoek.

De kosten van een goed functionerend chatprogramma staan tegenover de huurkosten van een accommodatie ten behoeve van traditioneel kwalitatief marktonderzoek.

Nucifora (2000) gaf tevens aan dat de beloningen voor deelnemers lager zouden zijn. Gegeven het feit dat online deelnemers per saldo minder tijd kwijt zijn aan het onderzoeksproces, en daarnaast bij het Online Forum ook nog eens zelf het tijdstip van deelname kunnen bepalen, is er een basis voor lagere respondentvergoedingen.

6.4.2. Ontbreken geografische barrières

Het ontbreken van geografische barrières zou een ander voordeel van online kwalitatief onderzoek kunnen zijn (Edmunds, 1999). Op basis van onze bevindingen kan

alleen aangetoond worden dat het bij online onderzoek relatief eenvoudig is om deelnemers uit verschillende regio's in Nederland te werven terwijl wervingsbureaus vaak regiogebonden zijn.

6.4.3. Dominantie, anonimiteit, en zelfpositionering

In een online setting zou volgens Aoki (1999) de kans relatief gering zijn dat enkele deelnemers de discussie zouden domineren. Hierdoor zouden de deelnemers vaker gelijkwaardige gesprekspartners zijn. Hieraan zou ook de grotere mate van anonimiteit (Stewart, 1998) kunnen bijdragen. Bij de Online Group blijkt de inbreng van de deelnemers, zowel kwantitatief als kwalitatief, redelijk aan elkaar gelijk te zijn. Dit komt overeen met bevindingen van Boetcher (2000). Grotere verschillen treden in dit verband op in de Online Fora. Deelnemers die relatief veel tijd wensen te besteden aan het Online Forum kunnen namelijk zelf bepalen hoe vaak ze wensen in te loggen. Dit zou er toe kunnen leiden dat deze deelnemers zich in elke interactie mengen, en daardoor als een soort stoorzender gaan optreden. Met betrekking tot anonimiteit valt op dat de onderzoeksdeelnemers met name tijdens de Online Fora vaak informatie uit hun werksfeer en hun persoonlijke levenssfeer prijsgeven. Dit kan worden opgevat als een uiting van de eerder beschreven behoefte om zichzelf te positioneren ten aanzien van de overige deelnemers.

6.4.4. Unieke informatie

Zowel de Online Group als het Online Forum produceren meer unieke informatie. Dit komt overeen met bevindingen van Parent (2000). Uit eerdere onderzoeken bleek al dat deelnemers in online discussies eerder geneigd zijn tegen meningen van anderen in te gaan. In dit onderzoek bleek dit met name bij het Online Forum het geval.

6.4.5. Non-verbale elementen en groepsdynamiek

Het ontbreken van non-verbale elementen zou volgens Greenbaum (1998) een van de grootste nadelen zijn van online groepsdiscussie. In de praktijk bleek inderdaad dat door dit manco het voor een moderator ten eerste erg moeilijk was om overzicht te houden. Ten tweede bleek het soms ook moeilijk om de verkregen informatie op een juiste manier te interpreteren omdat hierbij non-verbale indicatoren ontbraken. Dat hierdoor, zoals Greenbaum veronderstelt, van groepsdynamiek geen sprake zou kunnen zijn is in deze studie in het geheel niet gebleken.

6.4.6. Extreme standpunten

Volgens Silverman en Zukergood (2000) zou een online setting ertoe leiden dat deelnemers gemakkelijker extreme standpunten zouden innemen, en eveneens eerder andere deelnemers zouden aanvallen (zie Bellamy & Hanewicz, 1999). Geen van deze beide ongewenste effecten kan worden vastgesteld. Deelnemers van met name het Online Forum gaven weliswaar enkele malen aan het niet met elkaar eens te zijn maar dit gebeurde voortdurend op een correcte wijze. Een mogelijke verklaring hiervoor is dat Silverman en Zukergood gebruik maakten van natuurlijke, vrij toegankelijke online groepsdiscussies, in plaats van gemodereerde focusgroepen, met speciaal geselecteerde en uitgenodigde respondenten.

6.4.7. Onderwerpgerichtheid

Een mogelijk nadeel aan online groepsdiscussies zou volgens Tse (1999) zijn dat deel-

nemers snel geneigd zouden zijn om af te dwalen van het te bespreken onderwerp. Het onderzoek geeft aan dat een goede instructie, zorgvuldig vooraf geformuleerde vraagstellingen, en een attente houding van de moderator dit probleem kan voorkomen.

6.4.8. Type vraagstelling (stellingen vs. open vragen)

Het type vraagstellingen dat zich zou lenen voor online groepsdiscussie zou beperkingen kunnen opleveren. Zo zouden 'open' vragen leiden tot het eerder genoemde afdwalen van respondenten naar andere onderwerpen. Een remedie hiertegen is het gebruik van stellingen aan het begin van de sessie. Dit leidt echter tot minder doorvraagmogelijkheden in het Online Forum. In de online chat-groepsdiscussie blijkt de hoeveelheid informatie verkregen naar aanleiding van de stellingen beduidend minder te zijn dan bij Online Fora. Hierbij is doorvragen derhalve gewenst.

6.4.9. Software ten behoeve van online onderzoek

Als mogelijk nadeel noemt Levy (1998) onder meer de software-problematiek. De software dient betrouwbaar en gebruiksvriendelijk te zijn. Dit geldt ook voor de netwerkverbindingen. Het ontwerp dient zodanig te zijn dat het zowel de moderator als de deelnemers helpt overzicht van het proces te houden, en de deelnemers helpt en ondersteunt bij hun reacties. Tenslotte dient het uiterlijk een bijdrage te leveren aan de algemene motivatie, en het serieuze en professionele karakter van het onderzoek.

6.5. Aanbevelingen

Het geheel van de onderzoeksuitkomsten overziende kunnen de volgende aanbevelingen worden geformuleerd. Eerst zullen de aanbevelingen voor de methodiek van de Online Group worden gepresenteerd. Deze worden gevolgd door die van het Online Forum.

6.5.1. Online Group

Bij voorkeur dienen twee moderators de sessie te leiden. De taakverdeling is dan als volgt. Eén moderator dient zich met name te concentreren op de inhoudelijke interactie met de deelnemers, de andere moderator dient zich meer met de procesmatige aspecten bezig te houden.

Een centrale taak hierbij is het geven van persoonlijke en centrale feedback. Hier dient de moderator voortdurend afwegingen te maken: te snel reageren naar een individuele deelnemer kan leiden tot teveel 'ondersnjes' tussen de moderator en een deelnemer, en kan ook tot gevolg hebben dat een deelnemer voortijdig de gelegenheid wordt ontgenomen om het antwoord af te maken.

Een tweede belangrijke taak betreft het geven van leespauses aan de respondenten. Hiermee verkrijgt de moderator meer controle over het communicatieproces, aangezien dan duidelijk is waarmee een respondent doende is. Dit maakt het ook gemakkelijker om op de juiste momenten door te vragen, naar aanleiding van een initieel antwoord van een respondent.

De sessie dient aan te vangen met een algemene instructie. Hierin komen onder meer de duur, de opbouw van de sessie, en de 'gedragscodes' aan de orde. Tevens is dit de fase waarin mogelijke 'antwoordcodes' met de respondent worden doorgenomen, of

geoefend. Het gaat hier om, bijvoorbeeld, het gebruik van ‘emoticons’ (Harris, 1996).

Het te gebruiken softwareprogramma dient over functies te beschikken waarmee overzicht en controle wordt gecreëerd voor de moderator(s). Voorbeelden zijn een functionele kleurindeling van het scherm, waardoor onderscheid kan worden gemaakt tussen achtereenvolgens de moderator, een door de moderator geadresseerde respondent, en de overige respondenten. De te stellen vragen dienen vooraf in het programma ingevoerd te kunnen worden, teneinde de snelheid te kunnen opvoeren waarmee deze in de sessie kunnen worden ingebracht.

Een deelnemersaantal van 3-5 personen lijkt op dit moment het optimale aantal deelnemers voor een Online Group.

6.5.2. Online Forum

Ook hier dient de moderator bewuste en onderbouwde keuzes te maken wat betreft persoonsgebonden versus algemene feedback. Overwegingen hierbij zijn dat persoonsgebonden feedback individueel motiverend werkt, dit vanwege de zichtbaarheid van de persoonlijke inbreng, en omdat de respondenten bij deze werkwijze een groeps(leden)gevoel ervaren. Dit heeft gunstige effecten op de productiviteit van de forumleden.

In het algemeen kan worden gesteld dat de effectieve moderator een attente en actieve houding uitstraalt naar de deelnemers, en dat contacten tussen moderator en deelnemers altijd een persoonlijk element hebben.

Onderscheid kan worden gemaakt tussen een actief en een passief forum. In een passief forum krijgen de deelnemers eenmaal, of een gering aantal keren, feedback van de moderator, in de vorm van een samenvatting van de reacties van de deelnemers in een bepaalde ronde. Daarna hebben de leden de vrijheid om al dan niet te reageren, en worden niet meer aangestuurd. Tegenover het voordeel van een meer ‘natuurlijke’ dialoog kunnen als nadelen worden genoemd het optreden van minder interacties, meer uitval van de deelnemers, en verhoudingsgewijs minder bruikbare antwoorden. Gemodererde, actieve fora verdienen om deze redenen de voorkeur boven gemodererde, passieve fora, en ‘natuurlijke’ fora.

De moderator dient zich in het algemeen zichtbaar in de discussie te mengen, en wel op een interactieve, sturende wijze. Een groter aantal interacties tussen de deelnemers bevordert in het algemeen de kwaliteit en de kwantiteit van de gegevens. Doorvragen, of een toelichting vragen (‘proben’) leidt tot meer diepte, zoals dit bij de traditionele focusgroep het geval is. Een actievere moderator houdt de motivatie (taakgericht) beter en langer op peil, en verhoogt en verlengt eveneens de sociaal-emotionele gerichtheid van de deelnemers.

‘Onderonsjes’ kunnen worden vermeden door persoonlijke feedback van een algemeen element te voorzien.

Het management van de deelnemers vergt speciale aandacht. Actieve deelname dient te worden beloond. Hiertoe wordt het aantal keren dat een deelnemer inlogt geregistreerd.

Deelnemers kan een aparte ‘link’ worden aangeboden. Hiermee kan een deelnemer

contact opnemen met moderator buiten het forum om. Omgekeerd dient de moderator een attentiemail naar een deelnemer te sturen wanneer er op zijn of haar bericht is gereageerd.

Het softwareprogramma dient functionaliteiten te bezitten die leiden tot controlemogelijkheden voor de moderator, en tot overzicht en structuur voor de deelnemers. Zo moet het voor een deelnemer, door middel van kleurgebruik, zichtbaar zijn wanneer een bericht door deze deelnemer is gelezen.

Het minimum aantal deelnemers aan een Online Forum is 10 personen. Met dit aantal deelnemers is er nog sprake van een groepsgevoel, en groepsdynamiek. Dit blijkt onder meer uit het voorkomen van zogenaamde 'kettingreacties': het voorkomen van reactie op reactie door diverse respondenten, naar aanleiding van een initiële reactie van een deelnemer. Het aantal is dan nog klein genoeg om minimumvormen van beheersbaarheid te realiseren tijdens de sessie.

7. ONLINE KWALITATIEF ONDERZOEK: KENNISLEEMTES EN UITDAGINGEN

De uitdagingen waarvoor het online kwalitatief marktonderzoek zich gesteld ziet zijn talrijk, en groot in omvang. **Vooralsnog ontbreekt een breed, gedifferentieerd inzicht in, en een overzicht van de dimensies van het Internet-Umfeld.** Dit geldt ook voor de wijze waarop deze dimensies het voorkomen van bepaalde rolidentiteiten bevorderen, verwachtingen wekken bij respondenten, en de aard en de inhoud van de interacties bepalen. De wijze waarop het Umfeld ongewenste kleuringen teweegbrengt in de verzamelde gegevens is grotendeels nog onbekend. Hiermee verbonden is de vraag welk type informatie het Umfeld oproept, en welk type informatie het Umfeld onderdrukt of weg-filt. Uiteraard dienen deze inzichten nader te worden gedifferentieerd naar specifieke online onderzoeksmethodieken, zoals de **Online Single**, de **Online Group**, de **Online Letter**, en het **Online Forum**. Meer algemeen is het gewenst te bepalen welke soorten onderzoeksthema's meer, respectievelijk minder of niet geschikt zijn voor online kwalitatief onderzoek.

De ontwikkelde online onderzoeksmethoden dienen in algemene zin te worden gevalideerd, mede in relatie tot de bestaande kwalitatieve methoden. Hierbij kunnen op vruchtbare wijze vormen van **triangulatie** worden toegepast (Jones, 1999b), waarbij de onderzoeksuitkomsten afkomstig van verschillende, maar vergelijkbare gegevensbronnen en meetmethodieken met elkaar worden vergeleken. Op het niveau van conclusies en aanbevelingen zijn er signalen dat er sprake is van vergelijkbaarheid tussen face-to-face-onderzoek en online onderzoek (Scholl et al., 2001).

Fundamentele vragen blijven nog open aangaande de aard van de groepsprocessen en de groepsdynamiek in een online omgeving. Vooralsnog lijkt het gewettigd om de door Gordon & Langmaid (1988) beschreven fasering hierbij als uitgangspunt te nemen. Vergelijkend onderzoek dient hierop empirische antwoorden te geven.

Principieel dienen online sessies aangestuurd en gecontroleerd te worden door

een moderator. Voor de wijze waarop deze moderator taaksturing en sociaal-emotionele sturing in de virtuele omgeving realiseert bestaan nog geen wetenschappelijk onderbouwde modellen, theorieën, of protocollen. Nieuwe vormen en optimalisaties dienen hiervoor te worden ontwikkeld, getest, en gevalideerd. **Aandacht verdient ook de wijze waarop de moderator intervenueert, interacteert met, en feedback geeft aan respondenten tijdens de onderzoekssessies.** In dit kader verdient een systematische analyse van type vragen enerzijds, en de aard en inhoud van de gegevens die op basis hiervan zijn verkregen anderzijds, speciale aandacht.

8. ONLINE KWALITATIEF ONDERZOEK: KANSEN EN TOEPASSINGSMOGELIJKHEDEN

Het leidt geen twijfel dat het online kwalitatief onderzoek zich een eigen plaats zal veroveren tussen de gevestigde onderzoeksmethodieken. Er zijn argumenten om te veronderstellen dat op termijn het online kwalitatief onderzoek zich hierbij een zekere dominantie zal verwerven, tezamen met haar natuurlijke antipode, het kwantitatief online onderzoek. Hiervoor is een tweetal redenen te noemen. In de eerste plaats gebruiken we een historisch argument en een analogieredenering. **Waar het *face-to-face-onderzoek* ooit dominantie verloor aan het *telefonisch onderzoek* door de ontwikkeling en verbreiding van de telefonie, en het massale aantal telefoonaansluitingen, zal het telefonisch onderzoek op zijn beurt momentum verliezen aan het *online onderzoek*.** Dit wederom door een technologische revolutie, het Internet. Op dit moment is al meer dan de helft van de huishoudens in Nederland 'online'. Naar verwachting zal op overzienbare termijn dit aantal gestegen zijn tot een grote meerderheid van huishoudens in Nederland. Eenzelfde patroon is dan te verwachten.

Een tweede argument: marktonderzoek betreft in beginsel het verstrekken van gegevens door dataleveranciers aan een markt- of beleidsonderzoeksbureau, of –instelling. Dataleveranciers zijn in dit proces onder meer individuen in hun rol van consument, burger, of bedrijfsvertegenwoordiger. **Waar geldt dat het Internet, als medium en als *Umfeld*, in dwingende mate tot de habitat van deze dataleveranciers behoort ligt het in de rede om juist dit medium in te zetten als communicatievehikel bij de verzameling van deze gegevens.** Hier geldt het principe van de ecologische validiteit: houdingen, beelden, emoties, en gedrag van personen dienen bij voorkeur te worden gemeten in de setting waarin de leden van de populatie deze tentoon spreiden. Dit vergroot de kans dat steekproefbevindingen ook daadwerkelijk worden teruggevonden in de empirische werkelijkheid.

Een tweetal vormen van validiteit verdient, naast de reeds besproken vormen, in deze context speciale aandacht: constructvaliditeit en externe validiteit.

Kwalitatief marktonderzoek is traditioneel sterk ten aanzien van constructvaliditeit. In het kwantitatief onderzoek vindt de operationalisatie van begrippen als klantloyaliteit, kwaliteit of tevredenheid door middel van een – meestal kleine – set van items plaats. Waar de dimensionaliteit van deze begrippen vooraf onvoldoende vaststaat is het niet zeker of dergelijke itemsets in de praktijk daadwerkelijk de beoogde begrippen representeren. Kwalitatief onderzoek daarentegen houdt zich juist in belangrijke mate bezig met het inventariseren en benoemen van de dimensies van der-

gelijke begrippen, en kent hiermee een andere, en eigen onderzoeksdoelstelling. **Waar het online kwalitatief onderzoeksproces de werking heeft die in het onderzoek is beschreven zijn er geen a priori redenen om te veronderstellen dat online kwalitatief onderzoek tot geringere constructvaliditeit zou leiden.**

Externe validiteit heeft in de context van kwalitatief onderzoek betrekking op het detecteren en beschrijven van *alle* soorten houdingen, opinies, beelden, en gedragsvormen ten aanzien van een product, dienst of issue, die in de populatie voorkomen, los van onderliggende aantallen (Groenland, 2001). In die zin voldoet het kwalitatief onderzoek, door middel van een weloverwogen steekproefplan, aan algemeen te stellen wetenschappelijke eisen ten aanzien van externe validiteit. De mogelijkheden van online kwalitatief marktonderzoek ten aanzien van de steekproefopbouw zijn aanzienlijk groter dan die van het klassieke marktonderzoek. Immers: plaats (de onderzoekslocatie), en in zekere zin ook tijd (Online Forum) zijn aan minder praktische beperkingen onderhevig. **Online kwalitatief marktonderzoek zal daarom naar verwachting in de praktijk tot uitkomsten leiden met een adequaat niveau van externe validiteit.**

Ook met de introductie van online kwalitatieve onderzoeksmethoden blijft er een principieel verschil bestaan tussen kwalitatief en kwantitatief onderzoek. De notie van **‘incommensurabele paradigma’s’ (onderzoeksbenaderingen die van wezenlijk verschillende opvattingen ten aanzien van mens en wereld uitgaan)** (Swanborn, 1981) blijft geldig. Dit laat onverlet de vele mogelijkheden om onderzoeksbenaderingen van beide types vruchtbaar met elkaar te combineren.

Met de introductie en nadere ontwikkeling van het instrumentarium van het online kwalitatief onderzoek verkrijgt de kwalitatieve benadering nieuwe impulsen. Het uitgangspunt van kwalitatief onderzoek – het verkrijgen van diepere inzichten in de belewingswereld van individuen, en groepen van individuen, rondom producten en diensten – blijft onveranderd. **Ongekende kansen en mogelijkheden ontstaan echter voor het onderzoeken van sensitieve onderwerpen in een meer anonieme setting, en voor het benaderen van ‘low incidence’-steekproeven.** Dit uitbreidingspotentieel zal zonder twijfel de algemene acceptatie van kwalitatief onderzoek bij onderzoekers en opdrachtgevers uit kwantitatieve onderzoeksscholen en –richtingen doen toenemen. Ten aanzien van kwalitatief onderzoekers mogen in dit verband effecten met betrekking tot toenemende professionaliteit worden verwacht.

Gesteld werd dat bij alle methoden van datacollectie de **context waarin de gegevens worden verzameld, het *Umfeld***, tot methodegebonden vertekeningen van de data leidt. Een consequentie hiervan is dat klassieke en online kwalitatieve onderzoeksmethoden elk hun eigen merites en zwakten hebben. Een analyse van het type kennis en inzicht dat gezocht wordt (probleem-analyse), in combinatie met het type dataleverancier (respondent) dient de grondslag te vormen voor de te maken keuze tussen deze beide benaderingen.

Voor online kwalitatief onderzoek gelden dezelfde theoretische beoordelingscriteria als voor het klassieke kwalitatief onderzoek. **Het maken van beredeneerde keuzes in het onderzoeksontwerp, en het toepassen van transparante en controleerbare procedures, zoals beschreven in de ANNABEL-methodiek (Groenland, 2001) doen opgeld voor beide onderzoeksvormen.**

De hier gerapporteerde studie vormt slechts een eerste aanzet tot de accumulatie van wetenschappelijke kennis op deze terreinen. De bevindingen stemmen echter bepaald hoopvol. **Waar toegepast-wetenschappelijk onderzoek de basis vormt voor gehanteerde online dataverzamelmethodeken en methoden van analyse, zal het kwalitatief onderzoek ook op dit terrein uiteindelijk de plaats innemen die het, vanwege haar eigen performance, toekomt.**

LITERATUUR

- Aoki, K. (1999) Virtual Focus Groups: Opportunities and Limitations. Retrieved February 2001, from the world wide web: <http://people.bu.edu/kaoki/IFDTC.htm>
- Bellamy, Al & Hanewicz, Cheryl. (1999). Social Psychological Dimensions of Electronic Communication. *Electronic Journal of Sociology*(4),1.
- Benedikt, M., (1991). Cyberspace: some proposals. In M. Benedikt (Ed.), *Cyberspace: First steps* (pp. 119-124). Cambridge: MIT Press.
- Boetcher, S. (2000). Retrieved February 2001, from the world wide web: <http://www.workz.com/content/593.asp>
- Burke, R. Retrieved February 2001, from the world wide web: <http://universe.indiana.edu/clp/rf/focusg.htm#b11>
- Döring, N. (1997). Kommunikation im Internet: Neun theoretische Ansätze. In B. Batinic (Ed), *Internet für Psychologen*. Göttingen: Hogrefe-Verlag.
- Edmunds, H. (1999). *The focus group research handbook*. Lincolnwood, IL: NTC Business Books/Contemporary Publishing.
- Fernback, J. (1999). There is a There There. Notes toward a definition of Cybercommunity. In Jones, S. (Ed.), *Doing Internet research. Critical Issues and Methods for Examining the Net*. (pp. 203-220), Thousand Oaks, LA: Sage Publications.
- Gordon, W. & Langmaid, R. (1988) *Qualitative Market Research: A Practitioner's & Buyer's Guide*. London: Gower.
- Greenbaum, (1998). Internet focus groups are not focus groups- so don't call them that! Retrieved February 2001, from the world wide web: <Http://www.groupsplus.com/pages/qmrr0798.htm>
- Groenland, E.A.G. (2001) *Online kwalitatief marktonderzoek. Een deerne van vele zinnen*. Oratie. Breukelen: Universiteit Nyenrode.
- Harris, C. (1996). An Internet Education. A guide to doing Research on the Internet. Wadsworth Publishing Company.
- Jones, S. (1999) Preface. In Jones, S. (Ed.), *Doing internet research. Critical issues and methods for examining the net*. Thousand Oakes, LA: Sage Publications.
- Jones, Steve. (1999b). Studying the Net. Intricacies and Issues. In Jones, S. (Ed.) *Doing internet research. Critical issues and methods for examining the net*. Thousand Oaks: Sage Publications.
- Krekt, R. & J. Paarlberg & B. Schigt & K. Weima. E-commerce en reclamebranche. 'Boer pas op je kippen'. 1999. In: C. Molenaar (Red.). *Veranderingen door Internet. Branches in de toekomst*. Alphen aan den Rijn: Samsom.
- Levy, P. (1998). The city built on free rent. *Internet World*, 9 (1).
- Mann, Ch. & Stewart, F. (2000). *Internet communication and qualitative research. A handbook for researching online*. Thousand Oaks, London: Sage Publications.
- Maslow, A. (1954). *Motivation and Personality*. New York: Harper and Row.
- Nucifora, A (2000). Retrieved, February 2001 from the World wide web: <http://www.nucifora.com/art146.html>
- Parent, M et al. (2000). *Information & Management*, 38, 47-58.
- Scholl, N. & Mulders, S. & Drent, R. (2001). On-line kwalitatief marktonderzoek. 'Interviewing the world at a finger tip'. In: A.E. Bronner et al. (Eds). *Ontwikkelingen in het marktonderzoek. Jaarboek 2001*. Haarlem: De Vrieseborch.

- Sheth, Y. & Sisodia (1997). Consumer Behavior in the Future. In: R.A. Peterson (ed.) *Electronic marketing and the consumer*. Thousand Oaks: Sage Publications.
- Silverman, G. & Zukergood, E. (2000) MarketNavigation, Retrieved February 2001, from the world wide web: <http://www.mnav.com/cligd.htm>
- Stewart (1998). Using the internet in qualitative public health research: A comparison of Chinese and Australian young women's Perceptions of Tobacco Use. *Internet Journal of Health Promotion*. Retrieved February 2001, from the world wide web: <http://www.Monash.edu.au/health/IJHP/1998/12>
- Swanborn, P. G. (1981) *Methoden van Sociaal-Wetenschappelijk Onderzoek*. Meppel: Uitgeverij Boom.
- Tse, C.B. (1999) Conducting electronic focus group discussions among Chinese respondents. *Journal of the market research society*, 41(4), 407-415.
- Utz, S. (1999) Untersuchungsformen in MUDs. In: B. Batinic & A. Werner & L. Gräf & W. Bandilla (Eds.). *Online Research. Methoden, Anwendungen und Ergebnisse*. Reihe Internet und Psychologie. Neue Medien in der Psychologie, Göttingen: Hogrefe.

4. *Wie doen niet mee?*¹

Een analyse van nonrespons bij een onderzoek naar voorzieningengebruik

I.A.L. STOOP

SAMENVATTING

Het Sociaal en Cultureel Planbureau (SCP) probeert bij het Aanvullend Voorzieningengebruikonderzoek (AVO) een zo hoog mogelijke respons te realiseren. Hiertoe heeft het onderzoeksbureau, GfK PanelServices Benelux, een veelvoud van responsverhogende maatregelen ingezet. In 1995 is geprobeerd de verschillen tussen respondenten en nonrespondenten in kaart te brengen. Op grond van deze ervaringen is in 1999 het dataverzamelingsproces tot in details vastgelegd, zijn gegevens over respondenten en nonrespondenten verzameld via koppeling aan buurtgegevens en door observatie van interviewers, en zijn nonrespondenten herbenaderd in een vervolgonderzoek met een zeer hoge respons. Deze combinatie van gegevens maakt het mogelijk de vertekening als gevolg van nonrespons te achterhalen. De resultaten kunnen de veldwerkstrategie bij volgende peilingen effectiever maken en bijdragen aan de ontwikkeling van kennis over nonrespons in Nederland. In het verlengde hiervan wordt gepleit voor een programmatische en systematische aanpak van nonresponsonderzoek, waarbij marktonderzoekers, universitaire onderzoekers en het Centraal Bureau voor de Statistiek hun krachten kunnen bundelen.

1. SURVEYONDERZOEK IN CRISIS?

In de Jaarboeken van de MarktOnderzoekAssociatie wordt regelmatig verslag gedaan van ontwikkelingen rond nonrespons bij persoons- en gezinsenquête's. Een overzicht hiervan is te vinden in De Leeuw, 1997; latere aanvullingen zijn de Kruijk en Hermans, 1998; Pondman, 1999; Van der Kolk, 2001; Kalfs, 2001; Carton, 2001. De aandacht voor nonrespons in de jaarboeken lijkt stabiel, net als overigens de respons bij surveyonderzoek, die ondanks enkele gunstige uitzonderingen onverminderd laag blijft. Langzamerhand is er in Nederland sprake van een respectabele hoeveelheid onderzoek naar nonrespons. Desondanks lijkt het probleem nonrespons noch bij universiteiten, noch in de wereld van het marktonderzoek noch door de officiële statistiek als erg urgent te worden ervaren. Recente ontwikkelingen kunnen hier mogelijk verandering in brengen.

¹ De interviewers bij het AVO en het nonresponsonderzoek, en Frans Louwen en zijn collega's van GfK hebben door hun inzet, volharding en vertrouwen in het eindresultaat een zeer waardevol databestand tot stand gebracht. Tevens dank aan Edith de Leeuw en medewerkers van GfK en het CBS voor hun nuttige tips en commentaar op de tekst. De eindverantwoordelijkheid voor dit artikel en de erin verwoorde opinies komen geheel voor de rekening van de auteur.

In het methodologieonderwijs van de faculteiten sociale wetenschappen in Nederland wordt nauwelijks aandacht besteed aan de kwaliteit van surveyonderzoek in het algemeen en de consequenties van nonrespons in het bijzonder. De stichting Uw Mening Telt (UMT), waarin gebruikers en aanbieders van marktonderzoek zich verenigden, probeerde de attitude ten opzichte van marktonderzoek te verbeteren en de respons te verhogen, onder andere door radioreclame met als boodschap *Doe je voordeel met je oordeel*. In juni 2000 organiseerde UMT een drukbezocht symposium over nonrespons met lezingen van internationale en nationale deskundigen, dat zelfs de landelijke pers haalde (Van Delft, 2000). Inmiddels is UMT, samen met de Nederlandse Vereniging voor Marktonderzoek en Informatiemanagement (NVMI) en de Vereniging van Marktonderzoekbureaus (VMO), opgegaan in de MarktOnderzoekAssociatie (MOA). De MOA heeft de kwaliteit van markt- en beleidsonderzoek, zowel nationaal als internationaal, in de breedste zin van het woord bevorderen, ontwikkelen en stimuleren. Op de website van de MOA (<http://www.marktonderzoekassociatie.nl>) staat voorlopig nog nauwelijks inhoudelijke informatie over hoe men de kwaliteit wil bevorderen. Het woord nonrespons kwam half 2001 op de site niet voor.

Ook individuele marktonderzoeksbureaus profileren zich zelden op het terrein van de bestrijding van nonrespons. Uit offertes blijkt dat in het Nederlandse marktonderzoek geen standaarddefinities van nonrespons worden gehanteerd en geen minimum responsniveaus. Om nonrespons te vermijden en de kosten te verminderen zijn enkele bureaus overgegaan tot het construeren van (internet)panels bestaande uit coöperatieve respondenten. Hierbij wordt vaak veel moeite gedaan om een panel representatief te krijgen en te houden in termen van socio-demografische en socio-economische kenmerken en kan men tevens een hoge respons garanderen. Het verschijnsel nonrespons (en de mogelijke vertekening) wordt zo verschoven naar de voorfase van de panelconstructie en daarmee ook minder zichtbaar.

Het Centraal Bureau voor de Statistiek (CBS) worstelt al jaren met een hoge nonrespons bij persoons- en gezinsenquête. Responspercentages rond de 55% zijn standaard. Over de vertekening die hierdoor kan ontstaan, is weinig literatuur beschikbaar. Door een intensievere begeleiding en aansturing van de interviewers en door reorganisatie van het veldwerkapparaat hoopt het CBS het tij te keren. Bij deze pogingen de hoge nonrespons terug te dringen worden, in een strategisch programma, ook universiteiten en marktonderzoekers betrokken. Daarnaast richt het CBS zich vooral op een andere strategie om het verschijnsel nonrespons minder relevant te maken. Het CBS probeert zoveel mogelijk vragenblokken uit enquêtes te vervangen door informatie uit registraties en administraties. Op deze manier vermindert de enquêtedruk, neemt het aantal overbodige vragen in surveys af, kunnen respondenten gericht worden benaderd en wordt het mogelijk om met een kleiner en beter interviewcorps de kwaliteit van surveys te verbeteren. Echter, lang niet alle benodigde informatie kan uit registraties worden gehaald. Woonwensen, politieke interesse, tijdbesteding, verplaatsingsgedrag, attitudes, meningen, waarden en de behoefte aan voorzieningen kunnen alleen in surveys worden gemeten.

Alle veldwerkorganisaties hebben te maken met een vlottend interviewcorps. Door de krappe arbeidsmarkt is het moeilijk om interviewers te krijgen en te behouden (De Heer, 1999). De interviewers, die toch een cruciale schakel zijn bij het verkrijgen van een hoge respons, verdienen met enquêteren ongeveer net zoveel als een werkster

(Survey fieldworkers, those forgotten plebeians who actually perform interviews... (Goyder, 1987, p. 30)). De kosten van veldwerk zijn onlangs aanzienlijk verhoogd, omdat veldwerkorganisaties met ingang van 2001 verplicht zijn om interviewers in dienst te nemen en premies voor hen af te dragen. Ook voor opdrachtgevers zijn de kosten van veldwerk een last. Zij hechten vaak meer waarde aan het gerealiseerde aantal interviews en aan een snelle oplevering dan aan de representativiteit van de gegevens. Het lijkt erop dat iedereen er bij gebaat is de mogelijk nadelige invloed van non-respons te minimaliseren.

Een parallel met recente ontwikkelingen in de Nederlandse veeteelt dringt zich op. Terwijl de consument naar de kiloknaller gaat voor de aanschaf van karbonaadjes en half-om-halfgehakt, maken de gekkekoeienziekte, varkenspest en mond- en klauwzeer duidelijk dat de intensieve veeteelt niet op de huidige wijze door kan gaan. Intensieve veeteelt en grootschalig surveyonderzoek behoeven een kwaliteitsimpuls, waar mogelijk gebaseerd op best practices. De beginnende samenwerking tussen CBS, MOA en universiteiten is een positieve ontwikkeling. En er zijn andere hoopgevende voorbeelden. Kalfs (2001) en Carton (2001) geven concrete invullingen van de manier waarop surveyonderzoek tot hoge respons en goede kwaliteit kan leiden. EUROSTAT en Europese statistische bureaus, waaronder het CBS, zijn druk bezig met de ontwikkeling van kwaliteitscriteria voor de officiële statistiek en voor surveys. Op initiatief van de European Science Foundation (ESF) is de afgelopen jaren het European Social Survey (ESS) voorbereid, een onderzoek naar waarden, attitudes en hieraan gerelateerd gedrag. Inmiddels is subsidie ontvangen van de Europese Commissie en de ESF en worden in 2002 data verzameld voor het ESS in minimaal 15 Europese landen. Het lokale veldwerk wordt gefinancierd door de nationale wetenschappelijke organisaties (NWO in Nederland). Het ESS heeft niet alleen een inhoudelijke doelstelling, maar tracht ook lokale best practices cross-nationaal uniform te implementeren en een grote hoeveelheid methodologisch onderzoek genereren op het terrein van surveykwaliteit. Een kwaliteitseis van het ESS aan uitvoeringsorganisaties in de deelnemende landen is een minimum respons van 70%.

Goed surveyonderzoek is mogelijk en nodig. Het kost aandacht, tijd, zorg en geld, maar uiteindelijk hebben beleid en wetenschap meer aan minder gegevens van goede kwaliteit dan veel omvangrijke databestanden waarvan de kwaliteit dubieus is. In dit artikel wordt een voorbeeld gegeven van een survey waarbij op alle mogelijke manieren getracht is een hoge respons te realiseren en daarnaast extra informatie te krijgen om de invloed van non-respons te analyseren en waar mogelijk te corrigeren. Het onderzoek is uitgevoerd in opdracht van het Sociaal en Cultureel Planbureau en uitgevoerd door GfK PanelServices Benelux. Vooraf wordt een pleidooi gehouden voor een strategische en gezamenlijke aanpak van non-respons in Nederland.

2. ONDERZOEK NAAR NONRESPONS IN NEDERLAND

De nauwkeurigheid van de resultaten van survey-onderzoek wordt bedreigd door een aantal typen fouten: steekproeffouten (door het benaderen van slechts een klein deel van de populatie), dekkingsfouten (een deel van de populatie wordt helemaal niet benaderd, zoals de institutionele bevolking), meetfouten en fouten als gevolg van non-

respons. Deze fouten zijn nooit helemaal te vermijden, soms komen ze in combinatie voor en soms leidt het bestrijden van de ene fout tot het vergroten van een andere fout. In dit hoofdstuk wordt alleen ingegaan op fouten als gevolg van nonrespons en dan nog alleen in grootschalig face-to-face onderzoek bij surveys onder de gehele Nederlandse bevolking.

Nonrespons bij survey-onderzoek onder personen in Nederland is hoog, hoger dan in andere landen en hoger dan in het verleden (De Heer, 1999). Een ideaal klimaat voor nonresponsonderzoek, zou je verwachten, enerzijds omdat door de hoge nonrespons de kwaliteit van surveyonderzoek aan kritiek bloot staat en er aanzienlijke verbeteringen te boeken zouden moeten zijn, en anderzijds omdat voor diepgaand onderzoek een hoge nonrespons aantrekkelijk is. Zo baseren Groves en Couper (1998) hun analyse van nonresponsprocessen op databestanden met responspercentages tussen de 82 en 97%, waarbij van de bereikte personen tussen de 87 en 98% daadwerkelijk meedoet (coöperatiepercentage). In het laatste geval maken ze zich zorgen over de kans om statistisch significante verschillen te ontdekken (p. 62). Bij Nederlandse databestanden hoef je je weinig zorgen te maken om de 'te lage' nonrespons. Nonresponsonderzoek in Nederland kan dus profiteren van een grote hoeveelheid beschikbaar materiaal (veel nonrespons), is wetenschappelijk en beleidsmatig relevant en kan leiden tot indrukwekkende verbeteringen. Van 50 naar 70 procent respons is nog eens wat anders dan van 90 naar 94 procent, al betekent dit in beide gevallen dat de nonrespons met 40 procent afneemt.

En inderdaad, langzamerhand is er een respectabele hoeveelheid onderzoek gepubliceerd. Eerder is verwezen naar de aandacht voor nonrespons in recente MOA-jaarboeken en het UMT-symposium in 2000. In 2001 wijdde het Wetenschappelijk Statistisch Agentschap (WSA) een gebruikersoverleg aan nonrespons, later in hetzelfde jaar volgt de Vereniging voor Statistiek (VVS) met een symposium over incomplete data. De inbreng uit Nederland in de International Workshop on Household Survey Nonresponse, vanuit CBS en universiteiten is groot. Nederlandse onderzoekers hebben een goede naam op het terrein van cross-nationaal vergelijkend onderzoek naar nonrespons (De Leeuw en Hox, 1998; De Heer, 1999). Daarnaast is in de afgelopen jaren veel nonresponsonderzoek gedaan rond het Nationaal Kiezersonderzoek (NKO), mede ingegeven door felle discussies in de pers en zelfs vragen in de Tweede Kamer (zie ook Van der Kolk, 2001) en rond het Onderzoek Verplaatsingsgedrag (OVG), waar aanzienlijke kwaliteitsverbeteringen zijn gerealiseerd (Kalfs, 2001). Ook de nonrespons bij de Politiemonitor is uitgebreid geanalyseerd (Schoen et al. 2000).

Behalve acties gericht op specifieke surveys wordt onderzoek gedaan naar wegen en imputatie ter correctie van nonrespons, verschijnen analyses van methode-effecten en de consequenties van telefonisch enquêteren (Hox et al., 1998; Pondman, 1998; Rispens en Van Goor, 2000) en verslagen van kleinschalige experimenten en nuttige overzichtsartikelen (De Leeuw, 1997). In 1994 heeft de universiteit van Amsterdam in opdracht van de Sociaal-Wetenschappelijke Raad (SWR) een verkennend onderzoek uitgevoerd naar nonrespons bij surveyonderzoek door universitaire en para-universitaire instellingen (Kalfs en Kool, 1994). Ook in 1994 is door de werkgroep Dataverzameling van de Vereniging voor Beleidsonderzoek (VBO) een onderzoek gedaan waar naar responsgedrag werd gevraagd (VBO, 1996).

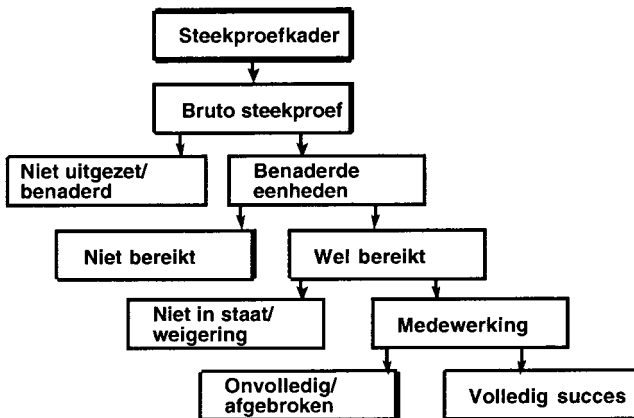
Initiatieven ter verbetering van de respons, zoals ook het geval bij het onderzoek waar de rest van dit hoofdstuk aan is gewijd, hebben vaak betrekking op één survey. Groves en Couper (1998) beschouwen deze geïsoleerde aanpak als een beperking bij het opbouwen van generieke kennis over nonrespons: *Given the wide variety of different survey designs and implementations, this makes it difficult to separate characteristics of households that may be associated with nonresponse from survey design features that may impact on the likelihood of making contact or gaining cooperation (p. 47)*. Ook Goyder (1987, p.82) pleit voor een systematische vergelijking van verschillende typen surveys om veronderstellingen over nonrespons te kunnen toetsen. Van dergelijk systematisch onderzoek zijn in Nederland nauwelijks voorbeelden te vinden. En ondanks de bovengenoemde literatuur blijft de nonrespons in Nederland, behoudens enkele uitzonderingen hoog. Symposia, experimenten en analyses van dataverzamelingsprocessen stralen zorg uit met soms kleine lichtpuntjes. Initiatieven bloeien op, leiden een geïsoleerd bestaan en verpieteren weer. Er lijkt een zekere matheid te ontstaan. Zoals De Kruijk en Hermans (1998) in een eerdere editie van dit jaarboek schreven: *(Markt)onderzoekers lijken wat weg te hebben van kledingverkopers. Als het gaat om non-response wordt dit afgedaan met: "dit is een gebruikelijk percentage weigeringen".... Nonresponse wordt blijkbaar als een gegeven beschouwd. (p. 55)*

Zoals algemeen bekend kunnen nadelen ook hun voordelen hebben. De sterk stijgende kosten van enquêtering en de twijfels aan de kwaliteit van surveys kunnen een katalyserende werking hebben, de matheid doorbreken en leiden tot een degelijk, integraal nationaal programma surveykwaliteit en nonrespons. Door de hoge nonrespons is de voedingsbodem zoals gezegd ideaal. Het CBS heeft de bestrijding van nonrespons en noncoverage bij de ontwikkeling van het nieuwe Meerjarenprogramma tot een strategisch project gemaakt en zoekt samenwerking met MOA en universiteiten (inclusief NWO en WSA). Naar aanleiding van de conferentie Survey Measurement and Process Quality die in 1995 in Bristol plaatsvond, pleitten Edith de Leeuw en Martin Collins (De Leeuw en Collins, 1995, pp. 214-217) al voor het vaststellen van een onderzoeksprogramma in internationaal verband. In Nederland kan men een voorbeeld nemen aan het geïntegreerde onderzoek naar surveykwaliteit in het Joint Program in Survey Methodology (JPSM) van de universiteiten van Maryland, Michigan en Westat (www.jpsm.umd.edu). In het verlengde kan ook in het onderwijs meer aandacht aan surveykwaliteit worden besteed. Waarom geen zomercursus surveyonderzoek?, analoog aan het Summer Institute van het Survey Research Center van het Institute for Social Research van de University of Michigan. Om het draagvlak te vergroten kan daarbij samenwerking worden gezocht met instellingen uit naburige landen die onderzoek doen naar surveykwaliteit en nonrespons, zoals ZUMA in Mannheim (www.gesis.org), het National Centre voor Social Research in Londen (www.natcen.ac.uk) en de universiteiten van Leuven en Essex.

In afwachting van strategische programma's wordt hier verslag gedaan van nonresponsonderzoek rond één survey, het Aanvullend Voorzieningengebruik Onderzoek van het SCP. Het onderzoek is bijzonder omdat op veel verschillende manieren informatie over nonrespondenten is verzameld. Daarmee kan het AVO-nonresponsonderzoek niet alleen de vertekening als gevolg van nonrespons bij een survey in kaart brengen, maar ook, hopelijk als 'best practice' dienen als startpunt voor nieuw onderzoek naar nonrespons.

3. NONRESPONS EN NONRESPONSONDERZOEK: DEFINITIES EN BRONNEN

Nonrespons bij face-to-face, schriftelijk en telefonisch onderzoek heeft verschillende oorzaken. Bij telefonisch onderzoek speelt de onbereikbaarheid van respondenten, door geheime nummers en het toenemende bezit van mobieltjes, een belangrijke rol. Daarnaast is het bij telefonisch enquêteren moeilijk om de benadering aan te passen aan de geselecteerde persoon. Hetzelfde geldt voor schriftelijk enquêteren. Een nadelige factor daarbij is dat de persoonlijke benadering ontbreekt. Bij telefonisch en face-to-face enquêteren kan de interviewer, afhankelijk van de aard van het contact en de respondent, proberen het ijs te breken. Als het gesprek eenmaal gaande is, komt de rest vanzelf. Enquêteren per internet of e-mail kent in Nederland bij algemene bevolkingsonderzoeken nog weinig toepassingen, met uitzondering van de computerpanels van het NIPO en Centerdata. Steeds meer komt men overigens tot de conclusie dat een meervoudige aanpak (*multi-mode*) vruchten af kan werpen, d.w.z. ofwel een combinatie van interviewwijzen, ofwel de respondent zelf de meest gewenste methode laten kiezen.



Figuur 1. Uitsplitsing respons en nonrespons.

Nonrespons is een complex verschijnsel. ‘Het’ nonresponspercentage van een survey zegt weinig tot niets, zolang niet bekend is ‘welk’ nonresponspercentage is gehanteerd en hoe de steekproefprocedure in zijn werk ging. In Nederland heeft de werkgroep Dataverzameling van de VBO (VBO, 1996) een aantal nationale en internationale definities van nonrespons bij elkaar gezet. In de VS geeft de AAPOR <http://www.aapor.org/ethics/responserates.html> een reeks definities, door het National Centre for Social Research in London en het Office for National Statistics (ONS) zijn deze geoperationaliseerd (Laiho et al. 2000; Lynn et al. 2000). Essentieel is dat nonrespons pas berekend kan worden als vooraf een bruto steekproef is gedefinieerd. Daarnaast is een onderscheid nodig tussen ongeldige eenheden (bijvoorbeeld leegstaande woningen bij een adressensteekproef), niet bereikte eenheden (nooit thuis) en personen die niet mee willen of kunnen doen. Voor de vergelijkbaarheid van surveys is een precieze operationalisatie van belang. Theoretisch is het onderscheid cruciaal tussen respon-

denten die niet bereikt konden worden en respondenten die wel zijn bereikt maar niet mee wilden of konden doen. Algemeen wordt aangenomen dat de vertekening door nonrespons bij deze groepen verschilt (Louwen, 1992). Het in dit hoofdstuk besproken onderzoek beoogt onder andere meer helderheid te verschaffen over deze verschillen.

Groves and Couper (1998) geven een overzicht van databronnen om informatie over nonrespons te verkrijgen. Hier worden ze vertaald naar de Nederlandse situatie. Een eerste belangrijke bron bevat gegevens over het steekproefkader. In Nederland is er bij steekproeven uit het Gemeenschappelijk Bevolkingsregister van respondenten en nonrespondenten een groot aantal gegevens beschikbaar over leeftijd, geslacht, huishoudenssamenstelling en nationaliteit. Helaas heeft alleen het CBS een rechtstreekse toegang tot dit steekproefkader. Bij een steekproef uit het PTT-afgiftepuntenbestand (de brievenbussen) kan met behulp van postcodegegevens een uitspraak gedaan worden over de woonbuurt van respondenten en nonrespondenten en de gemiddelde bevolkingssamenstelling (inkomen, opleiding, type huishouden). Gegevens over vier-cijferige postcodes uit de Geo-marktprofiel database spelen hier een nuttige rol.

Verwant hieraan zijn gegevens over respondenten (mits bereikt), de woning en de woonbuurt door de interviewer. Deze gegevens kunnen behulpzaam zijn gedurende het veldwerkproces, om de benadering van de respondent te 'tailoren', maar ook achteraf bij de analyse van nonrespons.

Een derde bron van informatie over nonrespondenten berust op de aannamen dat degenen die nooit thuis getroffen zijn in een survey, lijken op degenen die pas na veel bezoeken de deur open doen en dat degenen die weigeren lijken op de respondenten die slechts moeizaam kunnen worden overgehaald om mee te doen. Een dergelijke benadering vereist dat per steekproefeenheid en per bezoek nauwkeurig het resultaat wordt vastgelegd. Dat leidt tot de registratie van contactpatronen als in dit voorbeeld:

1^e interviewer: niet thuis; niet thuis; ongeschikt moment (zachte weigering); niet thuis; weigering;

2^e interviewer: niet thuis; afspraak; geslaagd gesprek.

Overigens worden de bovengenoemde aannamen in de literatuur volgens Groves & Couper (1998) niet altijd als realistisch ervaren. De houdbaarheid van de veronderstelling hangt mogelijk samen met het aantal bezoeken dat maximaal wordt afgelegd en de pogingen die worden gedaan om weerspannige respondenten toch over te halen. Aan de hand van de resultaten van de in dit hoofdstuk beschreven onderzoek, kunnen deze voorspellingen worden getoetst.

Een laatste, belangrijke bron van informatie over nonrespondenten is een vervolgonderzoek onder deze groep, ten minste op voorwaarde dat het merendeel van hen hier aan meewerkt. Samen met gegevens over alle benaderde respondenten (via databanken en observatie) en een gedetailleerde beschrijving van het veldwerkproces kan een vervolgonderzoek informatie leveren over nonresponsgedrag en kenmerken van nonrespondenten. Deze combinatie van bronnen wordt hierna beschreven.

4. SCP, AVO EN NONRESPONS

4.1. *Het SCP en het AVO*

Het Sociaal en Cultureel Planbureau doet sinds 1973 onderzoek naar de sociale en culturele situatie in Nederland. Het directe doel van dit onderzoek is het evalueren van beleid en het leveren van informatie aan departementen voor het ontwikkelen van nieuw beleid. Daarnaast worden de rapporten van het SCP veelvuldig gebruikt door gemeenten, instellingen, universiteiten en de pers. Voor het onderzoek wordt veelvuldig gebruik gemaakt van empirisch datamateriaal, deels afkomstig uit registraties en administraties, maar vooral uit surveys onder de Nederlandse bevolking. Een groot deel van deze surveys wordt gehouden door het Centraal Bureau voor de Statistiek. Daarnaast laat het SCP ook zelf surveydata verzamelen, alleen of in samenwerking met andere instellingen (zie voor een overzicht Stoop, 2000).

Het Aanvullend Voorzieningengebruikonderzoek (AVO) is een survey dat sinds 1979 om de vier jaar in opdracht van het SCP wordt gehouden (de term aanvullend betekent aanvullend op het werkprogramma van het CBS). Voor het onderzoek wordt een random steekproef getrokken van adressen uit het PTT-afgiftepuntenbestand. De uiteindelijk te realiseren respons bedraagt 6.500 huishoudens. Bij de eerste benadering wordt aan het hoofd van het huishouden of haar of zijn partner een mondelinge huishoudensvragenlijst voorgelegd. Vervolgens wordt voor ieder lid van het huishouden van zes jaar en ouder een schriftelijke vragenlijst achtergelaten. De vragenlijst voor volwassenen bevatte in 1999 bijna 300 vragen, waarbij een vraag een lijst van 27 activiteiten kan bevatten of 22 ziekten en aandoeningen of 11 verenigingen en organisaties. Voor jongeren is er een verkorte lijst die zo nodig ook door de ouders kan worden ingevuld. De interviewer haalt de schriftelijke lijsten na verloop van tijd weer op.

Het AVO gaat, zoals de naam al zegt, over het gebruik van en de behoefte aan een groot aantal (overheids)voorzieningen. Daarnaast wordt gevraagd naar een grote hoeveelheid achtergrondkenmerken:

- deelname aan sportieve, culturele en recreatieve activiteiten en het gebruik van voorzieningen op dit terrein;
- mediagebruik, vrijetijdsactiviteiten, hobby's;
- gebruik voorzieningen op het terrein van onderwijs, sociale zekerheid, rechtszekerheid, gezondheid, zorg en kinderopvang;
- inkomen, pensioen, uitkeringen.

4.2. *Nonrespons bij het AVO1995*

De respons bij het AVO daalde tot en met 1991 (in 1987 is de respons door een extra benadering en verlenging van de interviewperiode omhoog gegaan, daarna zette de dalende trend door). In 1995 heeft GfK na een Europese aanbesteding de opdracht gekregen voor het veldwerk. De hoogte van de respons was hierbij een belangrijk criterium. In 1995 heeft GfK 70% respons gerealiseerd; in 1999 (met meer inspanning en meer bezoeken dan in 1995) 66%.

Tabel 1. Respons AVO 1979-1999.

Jaar	Bruto steekproef (huishoudens)	Netto steekproef (huishoudens)	Netto steekproef (personen)	% respons (complete huishoudens)
1979	9.915	6.431	17.232	65
1983	9.908	5.774	14.869	58
1987	10.302	6.496	16.151	63
1991	12.797	5.458	13.105	43
1995	9.305	6.421	14.489	70
1999	9.300	6.125	13.490	66

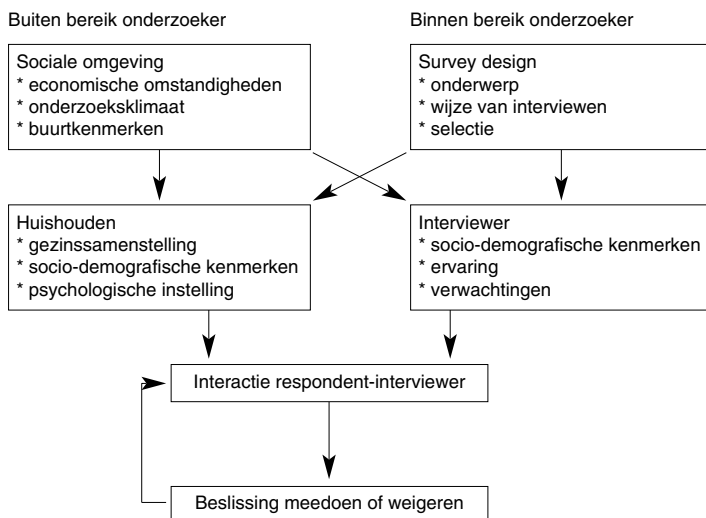
De maatregelen die GfK gebruikt heeft om de respons omhoog te krijgen van 43% naar 70% maken deel uit van een integraal systeem van benadering, supervisie en monitoring. Hiermee wordt ingespeeld op de verschillende factoren van surveyparticipatie uit het conceptuele model van Groves and Couper (1998):

- factoren buiten het bereik van de onderzoeker (sociale omgeving en gezinsstructuur); te denken valt aan het interviewklimaat en het hebben van kleine kinderen;
- factoren binnen het bereik van de onderzoeker (survey design en kenmerken interviewers); te denken valt aan het onderwerp en de lengte van de vragenlijst en de ervaring, betaling en motivatie van de interviewers, en
- factoren gerelateerd aan de interactie tussen interviewer en respondent; te denken valt aan het afstemmen van de benadering op de respondent ('leuke hond') of het opperen van de mogelijkheid op een geschikter moment terug te komen.

Bij het AVO kan, vanwege het longitudinale karakter van het survey de vragenlijst moeilijk worden aangepast. Een kortere vragenlijst had ongetwijfeld de respons verhoogd, maar had dan weer tot informatieverlies geleid. Hier lag een deel van het design dus buiten het bereik van de onderzoeker. Factoren waarop wel kon worden ingegrepen waren in 1995 andere designaspecten, de kwaliteit van de interviewers, een zorgvuldige monitoring en respect en aandacht voor de potentiële respondenten. Hieronder worden enkele aspecten van de benadering genoemd, met tussen haakjes een verwijzing naar elementen uit het model:

- de respondenten ontvingen een aanschrijfbrief met een informatiefolder (design, interactie);
- de aanschrijfbrief was waar mogelijk op naam gesteld (interactie);
- hooggekwalificeerde, goed geïnstrueerde interviewers werden ingezet, van wie de prestaties werden gemonitord (interviewer);
- per telefoon kon na een bezoek aan de deur zo nodig een afspraak worden gemaakt (design, interactie);
- de verdeling van de bezoeken naar uur van de dag en dag van de week werd zorgvuldig gemonitord (design, interviewer);
- de respondenten werden tot 15 keer bezocht op verschillende momenten van de dag (design);
- de reden voor weigering werd vastgelegd en na weigering werd de respondent waar mogelijk later benaderd door een nieuwe interviewer (design, interactie).

Centraal in de benadering van GfK staat respect voor de respondent (zie ook Dillman, 1999). Deze heeft recht op informatie over inhoud en doel van het onderzoek en dient zorgvuldig behandeld te worden. Afspraken dienen dan ook te worden nagekomen. De



Figuur 2. Factoren die respons beïnvloeden.

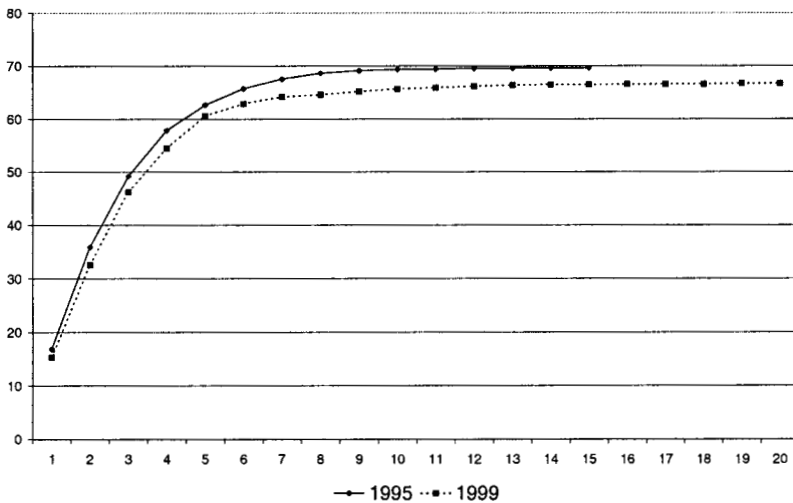
vele benaderingen en heruitzet van interviews waren essentieel voor het succes, maar waren alleen mogelijk omdat voor de dataverzameling een periode van vier maanden was gereserveerd.

Expliciet of impliciet blijken bij surveyonderzoek twee verschillen strategieën te worden gehanteerd (Groves and Couper, 1998). In de eerste strategie wordt het aantal participanten per tijdseenheid geoptimaliseerd. De interviewer besteedt zo weinig mogelijk tijd aan potentiële weigeraars of mensen die vaak niet thuis zijn. Het gaat erom zo snel mogelijk de vereiste 'n' binnen te halen. In de praktijk wordt deze strategie in het marktonderzoek veel gehanteerd (zie ook Stoop 1998). De aanpak van het AVO berustte op de tweede strategie, waar het aantal weigeraars of onbereikbaren wordt geminimaliseerd. Hier gaat het er om de eenmaal benaderde persoon over te halen mee te doen. Daarvoor moet de interviewer soms vaak terugkomen of zich tactisch terugtrekken als blijkt dat het de respondent even niet in de stemming is. Ook kan het nodig zijn op een later tijdstip een verse interviewer het nog eens te laten proberen. Bij deze tweede strategie wordt de kans op succes hoger naarmate de interviewer langer met een respondent aan de praat blijft.

Uit de tot dus ver verrichte analyses is niet duidelijk welk aspect van de veldwerkstrategie het meest effectief was. Wel blijkt (zie ook figuur 3) dat bij maximaal drie bezoeken de respons niet boven de 50% uitkomt. In 1991 deed een van de acht huishoudens niet mee omdat ze nooit thuis gevonden werden. In 1995 en 1999 kwam dit vrijwel niet voor. Het aantal wegeringen is in de loop van de tijd constant gebleven op ongeveer een kwart.

De responsgegevens uit het AVO1995 zijn geanalyseerd door Van Leest en Burhenne (1997) om na te gaan of er verschil is tussen makkelijke bereikbare respondenten, moeilijk bereikbare respondenten en (definitieve) weigeraars. Hierbij waren geen

gegevens per bezoek beschikbaar en was dus ook niet bekend welke respondenten na een initiële weigering alsnog waren overgehaald. In deze analyse werden aanvullende gegevens gebruikt uit postcodebestanden, populatiedata van CBS en uit de GfK-Mini-Census (GfK, 1999) en observatiegegevens van de interviewer. Uit de analyse bleek dat bewoners van (grote) steden, flatbewoners, 18-34 jarigen, alleenstaanden en kinderloze paren, hoog opgeleiden en de werkzame bevolking moeilijk bereikbaar waren. Het grote aantal contactpogingen (waar maximaal 3 in de Nederland gebruikelijk is) vergrootte de aanwezigheid van deze groepen in de steekproef, met uitzondering van grotestadsbewoners. Stadsbewoners, flatbewoners en ouderen weigerden relatief vaak mee te doen. Dit komt overeen met de resultaten van een literatuuroverzicht van Aarts en Van der Kolk (1999) die constateren dat ouderen, laagopgeleiden en stadsbewoners relatief vaak weigeren, terwijl jongeren, hoger opgeleiden, alleenstaanden en grotestadsbewoners vaker niet thuis worden getroffen. Overigens geeft het onderscheid in reden voor nonrespons tussen weigeraars en niet-bereikbaren ook aan dat het een gevaarlijke strategie is om alleen maar vaak langs te gaan en dus alleen het percentage niet-bereikbaren te reduceren, of alleen te proberen weigeraars over te halen. De structuur van de steekproef kan door een eenzijdige strategie juist verslechteren.



Figuur 3. Responsontwikkeling per bezoek bij AVO 1995 en AVO 1999.

In 1991 werden respondenten maximaal drie maal benaderd, in 1995 veel vaker. Dit verschil in design zou een effect kunnen hebben op de onderzoeksresultaten. Een andere reden om ‘vroeg’ met ‘late’ respondenten te vergelijken is de eerder genoemde veronderstelling dat ‘late’ respondenten het meest lijken op nonrespondenten. Bij het AVO1995 bestonden er grote verschillen in socio-demografische kenmerken en in voorzieningengebruik tussen beide groepen. De laatste verdwenen echter na weging met socio-demografische kenmerken. De nonresponsanalyse in 1995 wees erop dat het weinig zin leek te hebben om te streven naar een hoger responspercentage door het aantal contactpogingen nog meer te verhogen. Bovendien waren alle ‘onbereikbaren’ na 15 pogingen wel bereikt. Over het effect van het overhalen van weigeraars kon in

1995 geen uitspraak worden gedaan omdat slechts het totaal aantal contacten en het eindresultaat waren vastgelegd. Ook kon in 1995 niet worden onderzocht in hoeverre nonrespondenten op inhoudelijke kenmerken nu verschilden van nonrespondenten.

Waar na de analyses op het materiaal van 1995 behoefte aan bleek was aan precieze informatie over het resultaat van iedere contactpoging. Zo zijn ‘zachte’ weigeraars na verloop van tijd herbenaderd, maar is niet vastgelegd bij hoeveel personen dit succes heeft gehad. Alleen het eindresultaat van de reeks benaderingen is geregistreerd. Daarnaast was in 1995 geen informatie beschikbaar over inhoudelijke kenmerken van weigeraars (meer dan achtergrondgegevens, dus in dit geval over gegevens over voorzieningengebruik). Met beide typen gegevens zou het mogelijk moeten zijn om de responsgeneigdheid van verschillende groepen preciezer in kaart te brengen en om de vertekening door nonrespons te meten. Deze informatie kan dan weer worden gebruikt bij het wegen van resultaten en bij het design van de volgende peiling.

4.3. Nonrespons bij het AVO1999

In 1999 zijn respondenten door GfK op dezelfde wijze benaderd als in 1995. Ondanks een groter aantal benaderingspogingen en een verlengde veldwerkperiode kwam de respons in dit jaar niet boven de 66 procent (zie figuur 1). Vooral het opleveren van volledige sets voor alle leden van het huishouden bleek een probleem. In 1999 is uitgebreide informatie verzameld over het responsproces door het in extenso registreren van bezoekeninformatie en door het houden van een vervolgonderzoek. De volgende gegevens zijn beschikbaar om het responsgedrag verder te kunnen analyseren:

- per adres in de bruto steekproef
 - kenmerken van de buurt en de woning: bouwdatum, hoogbouw/laagbouw, onderhoudssituatie buurt en woning, binnenstad/buitenwijk; type woning; overige kenmerken woning;
 - type adres (woning, winkel, bedrijf); telefonische benadering vooraf; aantal huishoudens (bij contact); eindresultaat.
- per contactpoging
 - datum/tijdstip, resultaat;
- per contact (gesproken met respondent)
 - resultaat contact, eventueel reden weigering/niet in staat/afgebroken
 - leeftijd, geslacht
- per interviewer
 - leeftijd, geslacht

Op basis van het aantal contactpogingen, het aantal keren ‘niet thuis’ en het aantal en type weigeringen heeft GfK de respondenten met hiërarchische clusteranalyse ingedeeld in zes groepen (zie tabel 2). In de tabel kan men zien dat 17% van de respondenten in de loop van het onderzoeksproces minimaal een maal geweigerd heeft. Als men zich hierbij had neergelegd en deze weigeraars niet opnieuw had benaderd, zou de respons dus slechts 83% van 66% (dus 55%) bedragen. Eerder is al gezegd dat na drie pogingen ongeveer de helft van de respondenten ‘binnen’ is. Stoppen na drie pogingen betekent dus een respons van slechts 50% van 66% (dus 33%). Als men stopt na gemiddeld vier pogingen mist men volgens tabel 2 nog 12% van de potentiële respondenten, die overigens voor het merendeel ook niet zo simpel waren over te halen.

Overigens is er een complicatie bij de indeling van respondenten in makkelijk en moeilijk bereikbaren. Of iemand thuis wordt getroffen hangt immers niet alleen af van het feit hoe vaak men thuis is, maar ook van het moment waarop de interviewer aanbelt. Deze was geïnstrueerd de contactpogingen te spreiden over tijd van de dag en de dagen van de week. Als een respondent voltijd werkt en de eerste twee bezoeken vinden overdag plaats ('niet thuis'), is de respondent niet meer makkelijk bereikbaar. Zijn even bereidwillige mederespondent, bij wie de eerste benadering toevallig 's avonds plaatsvond, valt in tabel 2 wel onder de makkelijk bereikbaren. De registratie van tijdstippen van de bezoeken geven een indicatie dat dit effect enige invloed kan hebben: in totaal vond bijna een kwart van de eerste benaderingen 's ochtends plaats. Bij de makkelijk bereikbare en bereidwillige respondenten was dit 20%, bij de redelijk makkelijk bereikbare bereidwillige respondenten meer dan 30%.

De groep die makkelijk bereikbaar was en zonder problemen meedeed (I: toch meer dan de helft van de gerealiseerde steekproef) bestaat voor een groot deel uit gezinnen met oudere kinderen woonachtig in goed onderhouden koopwoningen in de provincie. Redelijk bereikbaar en makkelijk over te halen (II) was een groep respondenten uit stedelijke gebieden (veel Zuid-Holland), relatief vaak alleenstaand en wonend in flats met een inkomen tussen minimum en modaal. De kleine groep van moeilijk bereikbaren die toch zonder problemen meedeed (III) woont in steden in het westen, is alleen-

Tabel 2. Bereikbaar en bereid, AVO 1999.

		Bereikbaar			
		Makkelijk	Redelijk	Moeilijk	Totaal
Bereid	Makkelijk	I: 54% Gemiddeld twee bezoeken	II: 23% Gemiddeld vier bezoeken, waarbij twee maal niet thuis	III: 2% Gemiddeld negen bezoeken, waarbij acht maal niet thuis	79% Geen weigering
	Redelijk		IV: 4% Gemiddeld vier bezoeken, eenmaal niet thuis, twee maal andere reden		4% Geen weigering, andere redenen geen participatie
	Moeilijk		V: 7% Gemiddeld vier bezoeken, eenmaal weigert, eenmaal niet thuis	VI: 10% Gemiddeld zeven bezoeken, vijf keer niet thuis, een weigering	17% minimaal een weigering
	Totaal	54% Gemiddeld twee bezoeken	34% Gemiddeld vier bezoeken	12% Gemiddeld meer dan zeven bezoeken	100%=6125

staand of heeft kleine kinderen, verdient ook tussen minimum en modaal en heeft voor een groot deel een middelbare opleiding.

De middengroep, redelijk makkelijk bereikbaar en redelijk bereidwillig (IV) kwam voor een groot deel uit dorpen in het noorden van het land en woonde (dus) veelal in laagbouw. Het betrof hier voor een groot deel gezinnen met oudere kinderen.

De onwillige respondenten, die allemaal minimaal een keer geweigerd hebben, zijn onder te verdelen in een redelijk makkelijke en een moeilijk bereikbare groep. De eerste groep (V) bestaat uit voor een groot deel uit jonge huishoudens uit het zuiden van het land. De tweede, moeilijkste groep (VI) kent ook veel jonge huishoudens die echter vooral in niet al te dure huurwoningen in steden in het westen van het land wonen.

De zes groepen verschillen op veel van de onderzochte achtergrondvariabelen, voornamelijk gerelateerd aan woning, buurt, huishoudenssamenstelling, welstand en stedelijkheid en provincie. Wat betreft de mate van bereikbaarheid zijn er significante verschillen op al deze kenmerken. Voor de mate van bereidheid lijkt de welstand er niet toe te doen. GfK zal op een analoge manier als voor het AVO1995 onderzoeken of bereidheid en bereikbaarheid samenhangen met scores op doelvariabelen (voorzieningsgebruik). Net als in Van der Leest en Burhenne (1997) zullen daarbij systematisch verschillen tussen groepen in kaart worden gebracht, voor en na weging door achtergrondkenmerken.

Samengevat werd in 1995 een hoge respons gerealiseerd en waren er, uitgaande van het beperkte materiaal, geen aanwijzingen voor het optreden van bias voor nonrespons. In 1999 is daarom weer gestreefd naar een hoge respons, is het hele contactproces vastgelegd om onderscheid te kunnen maken naar problemen met bereikbaarheid en bereidheid, en is bovendien een vervolgonderzoek gehouden onder de harde weigeraars om zo echt te kunnen meten of bias optreedt als consequentie van nonrespons. In de rest van dit hoofdstuk staat het vervolgonderzoek onder weigeraars op de voorgrond.

5. VERVOLGONDERZOEK: DESIGN EN VELDWERK

Een vervolgonderzoek onder harde weigeraars kan de vertekening in kaart brengen als de respons binnen dat onderzoek hoog is. GfK en SCP besloten te streven naar een responspercentage van 80% voor dit onderzoek onder weigeraars. Om dit te realiseren waren maatregelen nodig op het terrein van design, interviewers, surveyklimaat, (inspelen op) huishoudenskenmerken en interactie, kortom op alle factoren in het conceptuele design van Groves and Couper (1998). Het onderzoek werd gehouden onder een random steekproef van 350 huishoudens die in het reguliere AVO minimaal twee keer geweigerd hadden. 21 Huishoudens vielen af wegens overlijden, verhuizen en administratieve vergissingen. Respondenten die verhuisd waren, werden niet op hun nieuwe adres herbenaderd, aangezien het oorspronkelijk om een adressensteekproef ging en woon- en buurtkenmerken een belangrijke rol spelen in de analyse.

In tegenstelling tot het hele AVO, waaraan alle personen van 6 jaar en ouder mee moesten doen, werd het vervolgonderzoek gehouden bij slechts een persoon per huis-

houden. Dit mocht degene zijn die de deur opendeed, mits het hier ging om het hoofd van het huishouden of zijn of haar partner. De vragenlijst was aanzienlijk ingekort (ca. 60 vragen over huishouden en persoon, interviewduur ca. 15 minuten) en bevatte achtergrondkenmerken, vragen naar voorzieningengebruik en vragen die ook in het AVO gesteld waren naar kenmerken die gerelateerd zijn aan responsgedrag (zie ook Groves en Couper, 1998, p. 131 e.v.) (kerkelijkheid, netwerken, vrijwilligerswerk, lid van verenigingen). Lange vragen met veel vervolgvragen werden soms ingekort en antwoordcategorieën gecombineerd. De afname was in principe mondeling, maar men kon de vragenlijst ook schriftelijk of telefonisch invullen en zelfs per internet.

Het beantwoorden van de vragenlijst was dus aanzienlijk minder belastend dan in het hoofdonderzoek. De wijzigingen in de aanpak, afname en inhoud van de vragenlijst kunnen natuurlijk effect hebben op het beantwoorden van de vragen. Om de vragenlijstverschillen bij de analyse als verklaring mee te kunnen nemen, is dezelfde vragenlijst ook voorgelegd aan een verse steekproef van 300 personen (de zogenaamde controlegroep). Zonder veel inspanning, met een gelimiteerd aantal bezoeken en zonder extra beloning voor de respondenten, werd hier een respons van 63% gerealiseerd, wat mogelijk verklaard kan worden door het feit dat dezelfde interviewers bij het vervolgonderzoek onder nonrespondenten en bij de controlegroep waren ingezet.

In het nonresponsonderzoek waren allerlei responsbevorderende maatregelen ingebouwd. Naast designfactoren als de korte vragenlijst, de beperking tot één persoon en de vrije keuze van interviewwijze, besteedde GfK speciale aandacht aan de interviewers, ondersteuning en advies, incentives, het belang van het onderzoek en het surveyklimaat.

- De interviewers waren speciaal geselecteerd omdat ze in het verleden bij het interviewen de beste resultaten behaalden. Dat werd hen ook verteld, en daardoor waren ze extra gemotiveerd.
- De interviewers verdiepten zich vooraf in de resultaten van de eerdere benaderingen van 'hun' weigeraars en hadden zo informatie over woning en buurt en reden van weigering.
- Het belang van het onderzoek kwam tot uitdrukking in de hoge beloning die de interviewers kregen en in een artikel in het NRC Handelsblad (Van Delft, 2000) dat vlak voor de start van het veldwerk verscheen en waarin het belang van het onderzoek werd benaderd. Daarnaast konden de interviewers een kaartje achterlaten met het directe telefoonnummer van de directeurs van GfK resp. SCP (*overigens heeft niemand ze gebeld*).
- Het veldwerk werd voorafgegaan door een uitgebreide briefing, waarin het doel en het belang van het onderzoek werden uiteengezet, waarin de beschikbare incentives werden gepresenteerd en waarin het vertrouwen werd uitgesproken dat de geselecteerde groep topinterviewers de target zou halen. Interviewers konden bij GfK terecht met vragen en problemen en stelden elkaar via een speciale nieuwsbrief op de hoogte van succesvolle strategieën.
- Voor respondenten die het vervelend vonden om 'lastig gevallen' te worden met interviews of direct mail (negatief surveyklimaat) was er een andere incentive om mee te doen. De interviewer legde uit hoe belangrijk het was om deze ene keer mee te doen om gegevens over nonrespondenten te verkrijgen, en kon bij medewerking beloven dat de geïnterviewde uit de databestanden van marktonderzoekbureaus en direct marketing organisaties werd gehaald.

- Tenslotte kregen de interviewers na de instructie per respondent cash 50 gulden mee. Dit bedrag mochten ze naar eigen inzicht besteden, op voorwaarde dat ze hun collega's informeerden over succesvolle strategieën. Tevens kregen ze een breed scala aan mogelijke beloningen voorgelegd (geld, cadeaubonnen, bloemen en loten, maar ook de mogelijkheid om bijvoorbeeld in een restaurant de vragenlijst te beantwoorden). Daarnaast konden ze publicaties van het SCP en GfK overhandigen.
- In dit geval werd niet vooraf een aanschrijfbrief aan de respondenten gestuurd, om ze niet af te schrikken. Wel hadden de interviewers een brief van de directeur van het SCP bij zich waarin het belang van het nonresponsonderzoek nog eens werd benadrukt.

Deze combinatie van maatregelen leidde niet tot de beoogde 80% respons, maar wel tot 70%, hetgeen hoog genoeg is om iets over weigeraars te kunnen zeggen. Let wel, het gaat hier om 70% van de hardnekkige weigeraars bij het hoofdonderzoek.

In het vervolgonderzoek is uitsluitend getracht een zo hoog mogelijke respons binnen te halen. Er is niet geëxperimenteerd met verschillende factoren, dus het is ook moeilijk om na te gaan wat nu het succes heeft bepaald. Drie bronnen kunnen hier enigszins uitsluitsel over geven. Ten eerste is aan de interviewers een variant voorgelegd van de vragenlijst van Snijkers et al. (1999). De resultaten hiervan zijn nog niet beschikbaar. Ten tweede hebben de interviewers het verloop van het veldwerk en hun acties geregistreerd. Daarnaast zijn alle interviewers uitgenodigd voor een focusgroepgesprek waarin systematisch gevraagd werd naar hun ervaringen in het veld (zie voor een verslag Stoop en Iedema, 2001 (<http://www.scp.nl/boeken/speciaal/spec6/nl/html/inhoud.htm>)).

Uit de registratie bleek dat interviewer hun eigen voorkeurstrategieën hadden en dat ze inventief en zuinig waren. Zo schrokken ze er niet voor terug om een interview via de intercom af te nemen, of een respondent, na overleg met een familielid, met een grote bos bloemen in het ziekenhuis te bezoeken en daar succesvol te interviewen. Soms kochten ze bloemen, maar dan meestal geen dure bossen. Sommige respondenten bleken overigens probleemloos mee te doen zonder dat enigerlei incentive werd ingezet. Van de geslaagde interviews vond 70% mondeling plaats, 20% telefonisch, 8% schriftelijk en 1% via het internet. Driekwart van de interviewers deelde een geldelijke beloning uit (meestal in de vorm van cadeaubonnen). Deze lag tussen de tien en tachtig gulden, met als gemiddelde ongeveer 35 gulden. Zelf vonden ze tien gulden eigenlijk net mooi voor zo'n kort interview (hetgeen aardig overeenkomt met de \$5 die Singer et al. (1999) effectief vinden). Slechts 10% van de respondenten maakte overigens gebruik van de gelegenheid om uit de databases van marktonderzoekers en telemarketeers te worden verwijderd (en één wilde er na het interview en de ontvangst van de beloning weer in).

Soms bleek men zoals gezegd zonder enig probleem bereid om mee te doen. Anderen waren volgens de interviewers met nog geen tien paarden over te halen. Van de weigeraars bij het vervolgonderzoek bleek overigens een vijfde bereid om bij een telefonische nabelronde toch nog de vragenlijst per telefoon in te vullen. Dit tot verbazing en soms tot frustratie van de interviewers, waarvan een enkeling alsnog bij de respondent verhaal ging halen (en een cadeautje brengen).

6. VERVOLGONDERZOEK: RESULTATEN

6.1. Samenstelling vervolgonderzoek en controlegroep

Voor de analyse van het vervolgonderzoek is een bestand geconstrueerd met gegevens van AVO-respondenten (maar dan slechts een persoon per huishouden, voor de vergelijkbaarheid), van de weigeraars met wie in het vervolgonderzoek een gesprek is gevoerd, en de controlegroep. De eerste analyse van de gegevens toont aan dat de steekproef van weigeraars een goede afspiegeling vormt van de totale groep nonrespondenten. Bij de samenstelling van de AVO-groep, de weigeraars en de controlegroep treden er verschillen op (zie tabel 3). Deze verschillen zeggen enerzijds iets over responsgeneigdheid, anderzijds kunnen ze mede een verklaring zijn van verschillen in doelvariabelen die optreden tussen respondenten en weigeraars. Bij de weigeraars zitten wat meer jongeren en ouderen, iets meer mannen en iets meer alleenstaanden. Het gaat hier dus steeds om degene met wie het gesprek gevoerd is; in het reguliere AVO ontloopt het aantal mannen en vrouwen elkaar niet veel omdat uiteindelijk alle personen in het huishouden zijn ondervraagd. Voor een goede analyse van compositie-effecten zal overigens naar meer variabelen moeten worden gekeken en naar bivariate effecten.

Tabel 3. Achtergrondkenmerken respondenten, weigeraars en controlegroep (%).

	AVO	Wei- geraars	controle		AVO	Wei- geraars	controle
leeftijd				omvang gezin			
<30	12	14	11	1	27	33	28
30-59	59	53	61	2	37	37	32
60+	29	31	28	3	13	11	17
				4	15	14	18
geslacht				5+	7	6	6
man	37	42	42				
vrouw	63	58	58	N	6125	233	176

6.2. Item-nonrespons en designeffecten

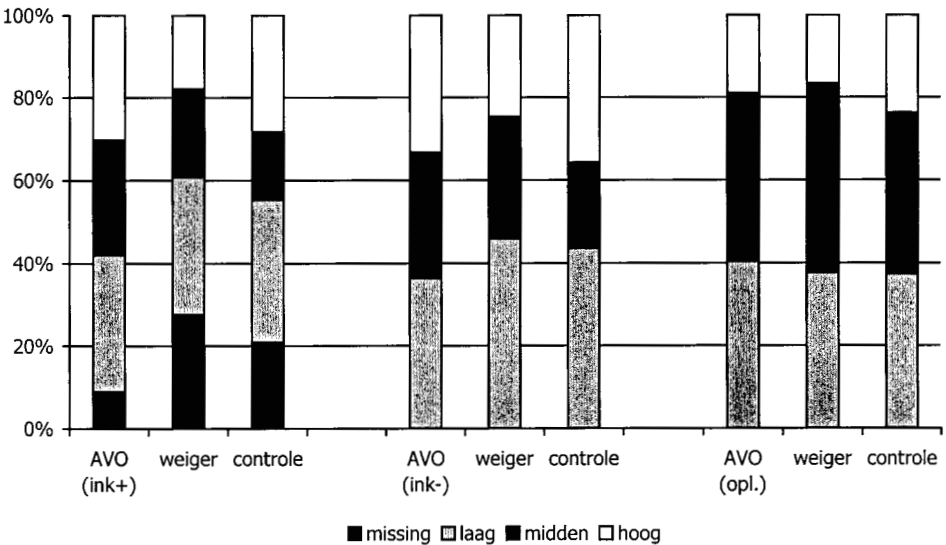
Bij het vergelijken van de antwoorden op inhoudelijke variabelen van respondenten en weigeraars is allereerst gekeken naar designeffecten. De resultaten van de controlegroep hielpen bij het opsporen hiervan. Als verrassend effect komt ten eerste naar voren dat de kwaliteit bij het nonresponsonderzoek beter is dan bij het reguliere AVO, althans als gekeken wordt naar item nonrespons. Dit is in strijd met de veronderstelling dat het pressen van respondenten tot medewerking leidt tot een slechts resultaat. Ook uit de gesprekken met interviewers is overigens in geen enkel opzicht gebleken dat de respondenten uit het vervolgonderzoek er met hun pet naar gooiden.

De meest plausibele verklaring voor het minder voorkomen van item nonrespons is dat het vervolgonderzoek grotendeels geheel mondeling werd afgenomen. De interviewer was waarschijnlijk nauwkeuriger dan de respondenten die bij het reguliere AVO

aan het eind van de lange vragenlijst wat vaker vragen naar lidmaatschappen van allerlei verenigingen en deelname aan allerlei activiteiten niet invulden. Bij de analyse van de reguliere AVO-gegevens is steeds aangenomen dat niet-invullen in deze gevallen 'nee' betekende.

Er is een uitzondering op de lagere item-nonrespons bij het vervolgonderzoek. In het reguliere AVO werd de interviewer op het hart gedrukt om item-nonrespons bij de vraag naar inkomen zoveel mogelijk te vermijden. Dit kenmerk is een kernvariabele bij onderzoek naar voorzieningengebruik. Bij het vervolgonderzoek gingen de interviewers veel soepeler te werk en benadrukten ze soms dat de respondenten niet het inkomen hoefden op te geven, hetgeen ze ook wat vaker deden (zie figuur 4). In dit geval lijkt er dus sprake van een trade-off tussen unit nonrespons en item nonrespons.

Uit de inspectie van designeffecten bleek ook dat bij de vraag naar kunstzinnige activiteiten de verkorte versie waarschijnlijk heeft geleid tot een iets lichtvaardiger opvatting van deelname aan zingen, fotograferen, textiele werkvormen etc. Hierop hoorde men alleen te scoren als men deze activiteiten in verenigingsverband beoefende of cursussen volgde. Dit kwam in de verkorte vragenlijst minder expliciet naar voren, vandaar de hogere participatie bij controlegroep en nonresponsgroep. Steun voor deze interpretatie geeft het feit dat er bij een vraag later in de vragenlijst naar lidmaatschap van een zang- of muziekvereniging veel kleinere verschillen zijn tussen beide vragenlijstversies. Bovendien is er in het reguliere AVO bij hetzelfde vragenblok tussen 1991 en 1995 ook sprake geweest van een trendbreuk met als meest plausible verklaring een wijziging in lay-out en het toevoegen van een vervolgvraag over financiële aspecten van deelname aan kunstzinnige activiteiten. Het betrokken vragenblok zal niet meer worden opgenomen in de analyse. Het nut van een controlegroep is hiermee evenwel overtuigend aangetoond.



Figuur 4. Inkomen (inclusief missing en exclusief missing) en opleiding van AVO-respondenten, weigeraars en controlegroep (ongewogen).

6.3. *Verschillen tussen respondenten en weigeraars*

De verschillen tussen respondenten en weigeraars zijn nog slechts oppervlakkig bekeken en niet gecorrigeerd voor compositie-effecten. Volgens figuur 2 zijn de weigeraars minder vaak hoog opgeleid en is het huishoudensinkomen lager. Figuur 2 geeft tevens het effect van de item-nonrespons aan. Een eerste inspectie van de verschillen tussen respondenten en weigeraars op voorzieningengebruik brengt een aantal verschillen aan het licht. Deze zijn voor het merendeel niet significant. De doelvariabelen waarvoor een significant verschil bestond tussen het reguliere AVO, de nonresponsgroep en de controlegroep zijn opgenomen in tabel 4. De resultaten zijn ongewogen weergegeven. Weging naar leeftijd en geslacht maakte overigens geen enkel verschil.

Uit tabel 4 blijkt dat de gezondheid van de weigeraars iets minder goed is dan die van de AVO-respondenten. Ze scoren ook minder hoog op de batterij vragen naar sportdeelname. Omdat het weinig zin leek te hebben om alle afzonderlijke sporten te vergelijken (voetbal, handbal, korfbal, hockey, etc.) is voor iedere respondent het aantal sporten geteld. Nu blijkt dat, waar van de respondenten 40% geen sport beoefent, dit geldt voor maar liefst de helft van de weigeraars. Het manco aan actieve sportdeelname wordt gecompenseerd door het feit dat de weigeraars vaker naar sport op tv kijken (35% van de AVO-respondenten kijkt nooit tegenover 21% van de weigeraars). De weigeraars hebben minder vaak een PC, lezen minder vaak boeken en gaan minder vaak naar concerten. De verschillen bij bezoeken van popconcerten, om een ander aspect van cultuurdeelname niet te verwaarlozen, worden voornamelijk veroorzaakt door het feit dat de deelname minder vaak 'onbekend' is. Als wordt aangenomen dat onbekend geen deelname betekent, gaan de weigeraars wat vaker naar popconcerten.

Uit deze eerste inspectie lijkt een voorzichtig beeld te rollen van weigeraars die iets minder participeren dan de rest van de bevolking. Deze veronderstelling wordt op losse schroeven gezet na enerzijds een vergelijking met de controlegroep en anderzijds na uitsplitsing naar moeilijke en makkelijke respondenten in het AVO. Zo lijken de weigeraars bij sport-kijkgedrag en boeken lezen zeer op de controlegroep en zijn de verschillen met de gewone respondenten groot. Dat zou wijzen op een vragenlijsteffect. Het blijft dan echter merkwaardig dat de weigeraars en de controlegroep zo op elkaar lijken, terwijl er tussen de subgroepen in het AVO wel degelijk verschillen bestaan. In andere gevallen lijken weigeraars en controlegroep weer helemaal niet op elkaar (bij sportbeoefening en PC-bezit) en vertoont de controlegroep grote overeenkomsten met het reguliere AVO. Ten slotte is het ook niet zo dat de moeilijkste respondenten in het gewone AVO nu zoveel op de weigeraars uit het vervolgonderzoek leiden. Dat is opvallend, omdat het in beide gevallen toch gaat om respondenten die minimaal een keer geweigerd hebben om mee te doen.

Samenvattend blijkt uit de eerste analyses van het vervolgonderzoek dat de verschillen tussen weigeraars en respondenten in het algemeen klein zijn (tabel 4 bevat de significante verschillen). Eenduidige uitspraken over effecten van het design (met behulp van de controlegroep) kunnen slechts in een enkel geval worden gedaan. De vergelijking tussen typen respondenten en weigeraars roept in eerste instantie ook meer vragen op dan dat er een duidelijk beeld uit voortvloeit. In dit geval geldt niet de stan-

Tabel 4. Verschillen tussen respondenten, nonrespondenten en controlegroep, ongewogen.

	Hoofdonderzoek AVO							Nonres- ponse- onder- zoek	Controle- groep
	Totaal	Bereikbaarheid							
		Makkelijk		Redelijk makkelijk		Moeilijk			
		Bereid							
	Makke- lijk	Makke- lijk	Rede- lijk	Moei- lijk	Makke- lijk	Moei- lijk			
gezondheid									
niet ingevuld	2	2	2	3	2	1	3		
zeer goed	23	22	24	17	24	21	26	20	12
goed	57	56	57	60	58	66	55	56	65
matig	15	15	14	16	15	11	14	18	18
slecht	2	2	1	3	1	1	1	5	5
zeer slecht	2	2	2	2	1	1	2	1	0
aantal sporten									
geen	40	41	40	37	39	32	34	50	42
1	23	24	21	26	26	33	24	27	26
2	17	16	18	23	17	19	18	11	17
3	10	9	9	8	12	6	13	6	5
4 of meer	10	9	12	6	7	10	12	6	10
sport op tv									
niet ingevuld	2	2	2	3	1	3	3	0	0
meer dan 1x per week	43	44	45	41	39	37	36	57	55
1 à 3 maal per maand	12	12	12	10	15	13	13	11	14
minder dan 1x per maand	8	8	7	9	10	8	10	10	10
nooit	35	34	34	36	34	39	38	21	21
bezit PC/laptop									
niet ingevuld	2	3	3	3	2	3	1		0
neen	42	43	40	41	37	31	40	56	42
ja	56	54	57	56	60	66	58	44	58
laatste boek gelezen									
niet ingevuld	2	2	1	2	1		1		1
> 6 maanden geleden	25	26	24	29	23	23	24	34	34
3 à 6 maanden geleden	5	5	5	5	6	4	4	7	7
2 à 3 maanden geleden	9	9	9	8	11	6	9	8	6
< 1 maand geleden	59	58	60	56	60	67	62	52	51
bezoek concert									
nooit	83,9	84,2	82,5	90,0	87,8	86,2	79,9	89,8	86,4
< 1x per jaar	1,0	0,8	1,7	0,4	0,7	0,7	1,0		0,0
1x per jaar	4,9	4,9	5,0	3,8	4,2	5,5	5,9	3,8	4,0
2 à 3x per jaar	6,4	6,1	6,8	5,0	5,2	7,6	8,4	3,8	4,5
4 à 11 x per jaar	2,7	2,8	3,1	0,4	1,5		3,7	2,6	4,0
meer dan 1x per maand	1,0	1,2	0,8	0,4	0,5		1,2		1,1
bezoek popconcert									
nooit	72,2	73,2	70,7	74,9	70,5	65,5	71,9	72,3	72,3
< 1x per jaar	5,8	5,7	5,5	5,0	7,2	8,3	6,0	6,4	7,3
1 x per jaar	8,3	8,3	9,0	5,4	9,4	9,7	7,4	10,6	9,0
2 à 3 maal per jaar	5,6	5,5	6,3	3,8	5,7	5,5	5,5	7,7	4,5
4 of meer maal per jaar	2,4	1,8	3,1	1,7	2,2	2,8	4,5	1,7	2,8
onbekend	5,6	5,8	5,6	9,2	5,0	8,3	4,9	1,3	4,0
N	6125	3338	1403	239	403	145	597	235	177

daarduitspraak dat meer onderzoek nodig is, maar dat verdere analyse van de overdaad aan datamateriaal dringend gewenst is.

7. EN WAT NU?

Uit het voorafgaande blijkt dat een hoge respons mogelijk is, ook bij belastende onderzoeken en ook in Nederland. Behalve allerlei concrete maatregelen lijken voor het bereiken van een hoge respons een aantal algemene, onderling samenhangende factoren van groot belang:

- respect voor de respondenten (Dillman 1999): geef informatie, leg uit waarom een onderzoek van belang is, hou je aan afspraken;
- respect voor de interviewers: leg uit waarom een onderzoek van belang is, betaal ze voldoende, geef ze middelen (tijd, geld, informatie) om een hoge respons te realiseren;
- motivatie van interviewers (zie ook Pondman, 1999);
- creativiteit van de interviewer in de fase waarin medewerking moet worden verkregen (zie ook Loosveldt en Carton, 2000; Goyder, 1987; Groves en Couper, 1998).

De geweldige hoeveelheid materiaal die is verzameld maakt een hele range van analyses mogelijk. Op basis van de gegevens uit het gecombineerde bestand kan onderzoek gedaan worden naar zulke uiteenlopende zaken als de responsgeneigdheid van verschillende groepen, de invloed van sociale isolatie en participatie, de effectiviteit van incentives en de mogelijkheid van alternatieve weegmethoden. Met de resultaten van de analyses moet het mogelijk zijn bij nieuwe peilingen gericht en evenwichtig energie te besteden aan het bereiken van moeilijk bereikbaren en het overhalen van weigerachtigen.

Zoals eerder in dit hoofdstuk gezegd is het klimaat voor nonresponsonderzoek in Nederland ideaal. CBS, universiteiten en marktonderzoekers hebben inmiddels de handen ineen geslagen en werken aan een strategisch aanpak van nonrespons. Eerder in dit hoofdstuk is gepleit voor geïntegreerd onderzoeksprogramma surveykwaliteit. Aanvullend hierop zouden de genoemde partijen in concreto kunnen bijdragen aan de verdere ontwikkeling van kennis over nonrespons door vergelijkbare data te verzamelen als hiervoor beschreven (uitgebreide registratie en vervolgonderzoek onder nonrespondenten) in aanvulling op een ander survey met andere typen vraagstellingen. Op deze manier kunnen de resultaten worden gegeneraliseerd en mogelijke vertekeningen als gevolg van nonrespons goed in kaart worden gebracht.

LITERATUUR

- Aarts, Kees & Van der Kolk, Henk (1999) Tegen de stroom in? Non-respons en Hawthorne effecten in het Nationaal Kiezersonderzoek. Paper gepresenteerd op de NVMC-voorjaarsconferentie 'De waan van de Dag', Utrecht, 23 April 1999.
- Bethlehem, J.P.G. and H.M.P. Kersten, H.M.P. (1984) Werken met non-respons. Voorburg: CBS
- Carton, A. (2001), Total Quality Management bij training en supervisie van interviewers. In: A.E. Bronner et al. *Ontwikkelingen in het marktonderzoek*. (pp. 105-121). Jaarboek 2001

- Marktonderzoek Associatie. Haarlem: De Vrieseborch.
- Delft, Dirk van, (2000) Liever niet. Sociaal en Cultureel Planbureau brengt enquête-weigeraars in kaart', NRC-Handelsblad, 10 juni 2000.
- De Heer, Wim (1999) International Response Trends, Results of an International Survey. *Journal of Official Statistics*, Vol. 15, No. 2, pp. 129-142 .
- De Heer, Wim (2000) Survey Practices in European Countries. In: Siobhán Carey (ed.). *Measuring Adult Literacy. The International Adult Literacy Survey (IALS) in the European context.* (pp 43-64). London: Office for National Statistics.
- De Leeuw, E.D. (1997), Nonrespons: Wat weten we en wat kunnen we eraan doen. In: A.E. Bronner et al. *Recente ontwikkelingen in het marktonderzoek.* (pp. 39-59). Jaarboek van de Nederlandse Vereniging voor Marktonderzoek en Informatiemanagement. Haarlem: De Vrieseborch.
- De Leeuw, Edith, and Martin Collins (1997) Data Collection and Survey Quality: An Overview. In: Lars Lyberg et al. *Survey Measurement and Process Quality.* New York: John Wiley & Sons, Inc.
- De Leeuw, E.D., en J.J. Hox (1998) Nonrespons in surveys: een overzicht. *Kwantitatieve Methoden*, 19, 31-53.
- Dillman, Don A. (1999) *Mail and Internet Surveys: The Tailored Design Method.* New York: Wiley.
- GfK (1999) GfK MiniCensus. <http://www.gfk.nl/data/cons scop/text/minicens.htm>.
- Goyder, John (1987), *The Silent Minority. Nonrespondents on Sample Surveys.* Cambridge: Polity Press.
- Groves, Robert M. and Couper, Mick P. (1998). *Nonresponse in Household Interview Surveys.* New York: Wiley.
- Hox, J., E. deLeeuw en G. Snijkers (1998) Fighting Non-Response in Telephone Interviews: Successful Interviewer Tactics. In: A. Koch and R. Porst, *Non-Response in Survey Research.* ZUMA-Nachrichten Spezial, no. 4, pp. 173-185. Mannheim: ZUMA..
- Kalfs, N. en E. Kool (1994) Ervaringen met nonrespons. Amsterdam: NIMMO.
- Kalfs, N. (2001) Van oude en nieuwe dataverzamelingstechnieken: onderzoek naar verplaatsingsgedrag. In: A.E. Bronner et al. *Ontwikkelingen in het marktonderzoek.* (pp. 89-104). Jaarboek 2001 Marktonderzoek Associatie. Haarlem: De Vrieseborch.
- Kolk, H. van der (2001) Meten wat we allemaal al weten. In: A.E. Bronner et al. *Ontwikkelingen in het marktonderzoek.* (pp. 151-168). Jaarboek 2001 Marktonderzoek Associatie. Haarlem: De Vrieseborch.
- Laiho, Johanna, Peter Lynn, Roeland Beerten and Jean Martin (2000) Alternative Definitions and Multiple Uses of Response Rates. Paper presented at the 11th International Workshop on Household Survey Nonresponse, 27-29 September 2000, Budapest, Hungary.
- Loosveldt, G., en A. Carton (2000) The impact of the interviewer contact strategies in face-to-face interviews on the final response rate and the data quality. Bijdrage aan het Non-response Symposium Amsterdam, 5 juni 2000.
- Louwen, F. (1992) Bereidwillig maar niet bereikbaar of bereikbaar maar niet bereidwillig. *Onderzoek*, juni 1992, 5-9.
- Lynn, Peter, Johanna Laiho, Roeland Beerten and Jean Martin (2000) A Project to Standardise Response Rate Estimation in the UK. Paper presented at the 11th International Workshop on Household Survey Nonresponse, 27-29 September 2000, Budapest, Hungary.
- Pondman, L.M. (1999) Overstag?! Reductie van het weigerpercentage in telefonisch onderzoek. In: A.E. Bronner et al. *Recente ontwikkelingen in het marktonderzoek.* (pp. 25-34). Jaarboek van de Nederlandse Vereniging voor Marktonderzoek en Informatiemanagement. Haarlem: De Vrieseborch.
- Rispens, Sonja, en Henk van Goor (2000) Een vertekend beeld van de maatschappij? Een onderzoek naar selectieve uitval in een telefonische enquête ten gevolge van onderdekking en nonrespons. *Sociologische Gids*, jaargang XLVII november-december, pp. 448-474.

- Schoen, E.D., P.R. Defize en M. Bakker (2000) Methodologische evaluatie van de Politie-monitor Bevolking. TNO-rapport FSP-RPT-000032. Delft: TNO-TPD.
- Singer, Eleanor, Robert M. Groves & Amy D. Corning (1999) Differential Incentives. Beliefs about practices, perceptions of equity, and effects on surver participation. *Public Opinion Quarterly*, Vol. 63, po. 251-260.
- Snijkers, Ger, Joop Hox and Edith de Leeuw (1999) Interviewers' Tactics for Fighting Survey Nonresponse. In: *Journal of Official Statistics*. Vol. 15, no. 2, pp. 185-198.
- Stoop, Ineke, en Jurjen Iedema (2001) Maar natuurlijk meneer, komt u binnen. In: *In plaats van kaarsen*. Nieuwjaarsboekje SCP 2001, pp. 44-49. Den Haag: Sociaal en Cultureel Planbureau.
- Stoop, Ineke, en Frans Louwen (2000) Interviewen is mensenwerk. *Facta*, november 2000, nr. 7, pp. 30-31.
- Stoop, Ineke (2000) Covering the Population, Sample Surveys at the Social and Cultural Planning Office. *Kwantitatieve Methoden*, jr. 21, nr. 65, pp.
- Van Leest, Joel, en Wim Burhenne (1997) Respons en nonrespons in het Aanvullend Voorzieningsgebruikonderzoek (AVO). Werkdocument 42, Rijswijk: Sociaal en Cultureel Planbureau.
- VBO (1996) Zicht op nonrespons? Werkdocument Werkgroep Dataverzameling van de Vereniging voor Beleidsonderzoek Ubbergen: Tandem Felix

5. *Kwaliteit van vragenlijsten bij survey- onderzoek onder etnische minderheden*

Effecten van de taal van het interview en de ethniciteit van de interviewer

H. VAN 'T LAND¹

SAMENVATTING

Dat het doen van onderzoek onder allochtonen geen gemakkelijke tak van sport is, wordt algemeen erkend. Maar vaak is nog onduidelijk hoe in de praktijk met methodologische knelpunten in minderhedenonderzoek omgegaan moet worden. In dit artikel wordt een tipje van de sluier opgelicht bij welk type survey vragen de kans op interviewer- en taaleffecten vooral te verwachten zijn. Ook wordt de kwaliteit van vragenlijsten onderzocht die in surveys gebruikt worden onder Marokkaanse respondenten. Afgesloten wordt met een aantal aanbevelingen hoe in de praktijk hiermee omgegaan moet worden.

1. INLEIDING

Begin jaren tachtig startte de Nederlandse overheid het minderhedenbeleid, dat specifiek erop gericht was om de ongelijkheid tussen allochtonen en autochtonen ten aanzien van arbeidsparticipatie, opleiding en inkomen te verkleinen (Minderhedennota; Ministerie van Binnenlandse Zaken). Maar nu, 20 jaar later, bevinden etnische minderheden (Turken, Marokkanen, Surinamers en Antillianen) zich nog steeds in een achterstandpositie als we kijken naar arbeidsparticipatie, opleiding en inkomen. Wat zijn nu de oorzaken hiervan? Waar ligt het nu aan? Moeten de oorzaken voor de achterstandsituatie gezocht worden in taalproblemen, cultuurproblemen, assimilatie- of integratieproblemen? Hiervoor is een belangrijke taak weggelegd voor het minderhedenonderzoek. Het spreekt voor zich dat onderzoek op dit vlak goed uitgevoerd moet worden. Er hangt immers nogal wat van af. Maar kunnen we het wel? Lukt het wel? Hebben we de goede middelen in handen? Trekken we wel de juiste conclusies? Dat het doen van onderzoek onder allochtonen geen gemakkelijke tak van sport is, wordt algemeen erkend (Liu, 1982; Choenni, 1984; Roelandt & Veenman, 1987; Bronner, 1988). Maar vaak is nog onduidelijk hoe in de praktijk met methodologische knelpunten omgegaan

¹ Dit artikel is gebaseerd op mijn proefschrift: Land, van 't, H. (2000). "*Similar Questions; Different Meanings. Differences in the meaning of constructs for Dutch and Moroccan respondents; effects of the ethnicity of the interviewer and language of the interview among first and second generation Moroccan respondents*"

moet worden. Meloen & Veenman (1990) hebben een eerste inventarisatie gemaakt van factoren die de kwaliteit van antwoorden van minderheden en de conclusies die op basis van dit onderzoek getrokken worden, negatief kunnen beïnvloeden. De auteurs onderscheiden onder andere de volgende factoren, die op basis van kwalitatieve groepsinterviews verzameld zijn:

- 1) de opdrachtgever van het onderzoek (*wie financiert het onderzoek?*)
- 2) het doel van het onderzoek (*wat wordt met de resultaten gedaan?*)
- 3) kenmerken van de respondent (*o.a. leeftijd, geslacht, etnische afkomst*)
- 4) kenmerken van de interviewer (*o.a. leeftijd, geslacht, etnische afkomst*)
- 5) kenmerken van de vragenlijst (*o.a. type vragen en vraagformat*)

In de survey praktijk is het usance om de etnische afkomst van interviewer en respondent te matchen, in de veronderstelling dat dit de kwaliteit van de antwoorden ten goede komt. Verondersteld wordt hoe 'Marokkaanser' de interviewsituatie is, hoe kwalitatief beter de antwoorden van Marokkaanse respondenten zullen zijn. Deze aanpak hoeft echter geenszins bij elk type survey vraag noodzakelijk te zijn. Voor de survey praktijk is inzicht op dit vlak goud waard aangezien het enorme consequenties met zich mee brengt. Als duidelijk wordt bij welke type survey vragen de kans op interviewer- en taaleffecten zo goed als afwezig zijn, hoeft er geen speciaal veldwerk apparaat opgestart te worden. In dat geval kan gebruik gemaakt worden van de oorspronkelijke vragenlijst en het vaste interviewcorps van een onderzoeksbureau. Dat brengt een enorme kostenbesparing met zich mee. Het onderzoek zal immers sneller uitgevoerd kunnen worden en de gegevens van Nederlandse en Marokkaanse respondenten zijn bovendien beter vergelijkbaar.

Dat de etniciteit van de interviewer ertoe doet, toonde Van Heelsum (1997) reeds aan. Zij liet zien dat de etniciteit van de interviewer een significante impact had op het type antwoord dat Surinaamse respondenten gaven op survey vragen. Zij vond dat de gemiddelde scores van respondenten die geïnterviewd werden door een Surinaamse interviewer systematisch verschilden van de antwoorden van respondenten die door een Nederlandse interviewer ondervraagd werden; dit was met name het geval bij een schaal die de oriëntatie van respondenten ten opzichte van contact met Surinamers meet.

In dit artikel zal ik een tipje van de sluier opgelichten² bij welk type survey vragen de kans op interviewer- en taaleffecten vooral te verwachten zijn. Ook onderzoek ik in hoeverre vragenlijsten, die ontwikkeld zijn voor Nederlandse respondenten, geschikte meetinstrumenten zijn voor Marokkaanse respondenten. De volgende onderzoeksvragen worden in dit artikel beantwoord:

Kwaliteit van vragenlijsten:

- 1) Zijn survey vragen die uitsluitend referen naar zaken die belangrijk zijn in de Nederlandse cultuur even belangrijk voor Marokkaanse als voor Nederlandse respondenten?
- 2) Neemt de geschiktheid van een vragenlijst toe voor Marokkaanse respondenten indien cultuurspecifieke survey vragen (vragen die referen naar zaken die belangrijk zijn in de Marokkaanse cultuur) aan een vragenlijst toegevoegd worden?

² Voor meer details over opzet en uitvoering van het onderzoek verwijs ik naar mijn proefschrift (zie Van 't Land, 2000).

Taal- en interviewer effecten:

3) Hebben de etnische afkomst van de interviewer en de taal die tijdens het interview gesproken wordt, invloed op de antwoorden die tweede generatie Marokkaanse respondenten geven op survey vragen?

2. METHODE

2.1. Kwaliteit van vragenlijsten en taal- en interviewer effecten

Om de kwaliteit van vragenlijsten in minderhedenonderzoek te onderzoeken, heb ik allereerst antwoorden van Marokkaanse en Nederlandse respondenten op een bestaande schaal geanalyseerd. Vervolgens heb ik een nieuwe vragenlijst ontwikkeld die rekening houdt met culturele verschillen tussen respondenten.

A) Analyse bestaande schaal: Voeg-schaal

Allereerst heb ik de validiteit vastgesteld van een bestaande schaal (de VOEG-schaal) voor Nederlandse en Marokkaanse respondenten. De VOEG-schaal is oorspronkelijk ontwikkeld door Dirken (1969) en meet psychobiologische stress en fysiek ongemak. De oorspronkelijke schaal bestaat uit 48 items. Zowel de betrouwbaarheid als de predictieve validiteit van de VOEG-schaal zijn hoog; de betrouwbaarheid is bepaald door de interne consistentie te berekenen en bedraagt 0.91, terwijl de predictieve validiteit varieert tussen 0.66 en 0.72 (Dirken, 1969). Hoewel de schaal ontwikkeld is voor Nederlandse respondenten, wordt deze tegenwoordig door de GG&GD en het CBS in grootschalige surveys gebruikt onder respondenten met verschillende etnische achtergronden. Het aantal items van de VOEG-schaal wordt in de praktijk teruggebracht. Zo reduceerde het Centraal Bureau voor de Statistiek (CBS) het oorspronkelijke aantal VOEG-items tot 13 items in de Nederlandse versie van de vragenlijst van het Leefsituatie-Onderzoek (LSO)³ en tot 6 items in de Marokkaanse versie van de vragenlijst. De GG&GD bracht de oorspronkelijke VOEG-schaal terug tot 21 VOEG-items. In een rapportage werd vermeld dat deze reductie geen effect had op de interne consistentie van de schaal (In 't Veld & Van Steenbergen, 1992).

In mijn onderzoek heb ik de ingekorte VOEG-schaal van de GG&GD gebruikt in de survey vragenlijst die afgenomen is onder Nederlandse en Marokkaanse respondenten. De validiteit van de VOEG-schaal heb ik bepaald met behulp van latente trek modellen. Het Rasch Model (Rasch, 1960) is een latente trek model dat bij uitstek geschikt is om de validiteit van een schaal te bepalen. Dit model gaat er namelijk van uit dat variaties in antwoorden van respondenten op VOEG-items uitsluitend bepaald worden door de scores van respondenten op de onderliggende trek (in dit geval 'psychobiologische stress en fysiek ongemak') die de VOEG-schaal meet. Het Rasch Model neemt verder aan dat de kans om een VOEG-item bevestigend te beantwoorden positief gerelateerd is aan de score op de onderliggende trek. Dus: hoe meer psychobiologische stress en fysiek ongemak een respondent ervaart, des te groter de kans dat een VOEG-item met 'ja' beantwoord wordt. Wel is het zo dat het Rasch Model gevoelig is voor afwijkingen in unidimensionaliteit van items. Daarom is voorafgaand aan de latente trek analyse, een HOMALS-analyse uitgevoerd teneinde een homogene set VOEG-items te identifi-

³ De LSO-vragenlijst werd tot 1993 gebruikt door het CBS.

ceren. Van de oorspronkelijke VOEG-schaal vormden uiteindelijk 8 VOEG-items een unidimensionale subschaal.

Geconcludeerd moet worden dat de validiteit van deze schaal niet in gevaar is als blijkt dat Marokkaanse en Nederlandse respondenten met dezelfde hoeveelheid psychobiologische stress en fysiek ongemak ook dezelfde kans hebben om de items bevestigend te beantwoorden. In dat geval verschilt de moeilijkheidsgraad van items niet significant tussen Nederlandse en Marokkaanse respondenten en is de schaal geschikt om te gebruiken in surveys onder Marokkaanse en Nederlandse respondenten.

Zodra echter blijkt dat de moeilijkheidsgraad van items significant verschilt tussen Nederlandse en Marokkaanse respondenten met dezelfde hoeveelheid psychobiologische stress en fysiek ongemak, komt de validiteit van een schaal in gevaar. In dat geval hebben Nederlandse en Marokkaanse respondenten, ongeacht hun gezondheidstoestand, een significant kleinere of grotere kans om een item bevestigend te beantwoorden. De schaal is in dat geval geen geschikt meetinstrument om te gebruiken.

B) Ontwikkeling van een nieuwe vragenlijst

Naast het analyseren van een bestaande schaal, is ook een geheel nieuwe vragenlijst ontwikkeld. Deze nieuwe vragenlijst houdt expliciet rekening met culturele verschillen tussen Nederlandse en Marokkaanse respondenten en bevat zowel vragen die relevant zijn in de Nederlandse cultuur als vragen die belangrijk zijn in de Marokkaanse cultuur. Nagegaan is of:

- a) vragen die relevant zijn in de Nederlandse cultuur even belangrijk zijn voor Marokkaanse en Nederlandse respondenten
- b) de geschiktheid van de vragenlijst toeneemt als cultuurspecifieke vragen toegevoegd worden.

Om de geschiktheid van de vragenlijst voor Marokkaanse respondenten te kunnen bepalen, zijn drie onderzoeksfasen uitgevoerd. Allereerst is in diepte-interviews nagegaan of, en zo ja welke, cross-culturele verschillen er bestaan in aspecten die Nederlanders en Marokkanen belangrijk vinden in 'werk', 'familie', 'vriendschap', 'opvoeding', 'relatie', 'huisarts' en 'sociale dienst'. Er zijn 5 diepte-interviews met Nederlandse en 2 interviews met Marokkaanse informanten gevoerd. Elk interview duurde 1.5 uur en werd op band opgenomen (zie ook van 't Land, 2000, hfd 3).

Vervolgens is in de tweede onderzoeksfase een pilotstudie uitgevoerd onder 52 Marokkaanse en Nederlandse respondenten. Deze pilot had twee doelen. Allereerst werd aan zowel de Nederlandse als de Marokkaanse respondenten gevraagd om aan te geven hoe belangrijk zij de 'Nederlandse' en 'Marokkaanse' aspecten vonden, die tijdens de diepte-interviews door de informanten genoemd werden. Hiertoe werden de aspecten die de informanten noemden in de vorm van stellingen op sorteerkarten gezet en de respondenten werd gevraagd om de aspecten op de sorteerkarten op belangrijkheid te sorteren. Ten tweede is in de pilotstudie een vragenlijst afgenomen die algemene gevoelens van (on)tevredenheid met de huisarts beoogt te meten. Deze vragenlijst is gebaseerd op 'Nederlandse' en 'Marokkaanse' aspecten die door de informanten genoemd zijn. Nagegaan werd welke survey vragen, die corresponderen met respectievelijk 'Nederlandse' en 'Marokkaanse' aspecten, het sterkst samenhangen met de algemene gevoelens van (on)tevredenheid van Marokkaanse en Nederlandse respondenten over de huisarts (voor details zie ook Van 't Land, 2000, hfd 3).

Tenslotte is in de derde en laatste onderzoeksfase een grootschalig survey uitgevoerd om empirisch vast te stellen of survey vragen die gebaseerd zijn op ‘Nederlandse’ aspecten even belangrijk zijn voor Nederlandse en Marokkaanse respondenten en of er in dit opzicht verschillen bestaan tussen eerste en tweede generatie Marokkaanse respondenten. In het survey zijn 578 mannelijke respondenten ondervraagd en er zijn over twee levensdomeinen (‘huisarts’ en ‘opvoeding’) vragen gesteld. De vragen zijn wederom gebaseerd op aspecten die door de informanten genoemd zijn tijdens de diepte-interviews. Wederom is vastgesteld of de geschiktheid van een vragenlijst toeneemt onder Marokkaanse respondenten indien cultuurspecifieke vragen (vragen die betrekking hebben op aspecten die belangrijk zijn in de Marokkaanse cultuur) aan een vragenlijst toegevoegd worden (zie Van ‘t Land, 2000, hfd 4 en 5). In deze derde fase is ook onderzocht welke invloed de etnische afkomst van de interviewer en de taal van het interview hebben op de antwoorden van tweede generatie Marokkaanse respondenten. Voor dit doel zijn enerzijds de resultaten geanalyseerd van een sorteertaak die respondenten uitvoerden op het levensdomein ‘opvoeding’ en anderzijds de antwoorden die de respondenten gaven op reguliere CBS-vragen, die in de vragenlijst waren opgenomen (zie ook Van ‘t Land, 2000, hfd 6).

2.2. Respondenten en interviewers

Om cross-culturele vergelijkingen te kunnen maken tussen Nederlandse en Marokkaanse respondenten, zijn uitsluitend Nederlandse en Marokkaanse respondenten geïnterviewd, die vergelijkbaar zijn wat betreft opleidingsniveau, sekse, woonplaats en leeftijd. In het onderzoek zijn uitsluitend *mannelijke* Nederlandse en Marokkaanse respondenten geïnterviewd, woonachtig in de randstad met een *opleidingsniveau* dat gelijk is aan MBO of lager.

Om onderscheid te kunnen maken tussen eerste en tweede generatie Marokkaanse respondenten, zijn *twee leeftijdscategorieën* gedefinieerd: een respondentgroep die 1) ouder is dan 29 jaar (eerste generatie Marokkaanse respondenten en vergelijkbare Nederlandse respondenten) en 2) een groep die ouder is dan 18 jaar en jonger dan 30 jaar (tweede generatie Marokkaanse respondenten en vergelijkbare Nederlandse respondenten). Het aantal jaren dat een respondent doorgebracht heeft op de basisschool, was het tweede selectie criterium waarmee vastgesteld werd of iemand tot de eerste of de tweede generatie behoorde. Tweede generatie Marokkaanse moesten vanaf hun vierde of vijfde in Nederland deelgenomen hebben aan het Nederlandse onderwijsstelsel om mee te doen aan het onderzoek. De *eticiteit* van Marokkaanse respondenten werd bepaald aan de hand van de geboorteplaats van beide ouders. Respondenten zijn van Marokkaanse afkomst als beide ouders geboren zijn in Marokko. De geboorteplaats van de respondent bleek hiervoor minder geschikt, aangezien het merendeel van de tweede generatie Marokkaanse respondenten geboren is in Nederland.

In de praktijk bleek het lastig om Nederlandse respondenten te werven met een opleidingsniveau dat overeenstemt met eerste generatie Marokkaanse respondenten. Het merendeel van de eerste generatie Marokkaanse respondenten blijkt namelijk geen opleiding te hebben gevolgd. Ook was het lastig om de gemiddelde leeftijd van eerste generatie Marokkaanse en vergelijkbare Nederlandse respondenten op elkaar af te stemmen; de eerste generatie Marokkaanse respondenten hebben een gemiddelde leeftijd van 35 terwijl de Nederlandse respondenten uit onze steekproef gemiddeld 45 jaar oud zijn. Teneinde het opleidingsniveau van eerste generatie Marokkaanse responden-

ten en Nederlandse respondenten zoveel mogelijk op elkaar af te stemmen, zijn in dit onderzoek uitsluitend Nederlandse respondenten geïnterviewd die een beroep uitoefenen waarvoor een opleiding niet direct noodzakelijk is (bijvoorbeeld ‘pakketjes transporteren’ of ‘post sorteren’). Het werven van tweede generatie Marokkaanse en overeenkomstige Nederlandse respondenten leverde geen problemen op.

In totaal werden 578 mannelijke respondenten geïnterviewd: 382 respondenten waren Marokkaans and 196 waren Nederlands. Van de Marokkaanse groep, behoorden 99 tot de eerste generatie generatie en 283 tot de tweede generatie Marokkkaanse respondenten. Alle respondenten werden geïnterviewd door mannelijke interviewers. De Nederlandse respondenten werden geïnterviewd door Nederlandse interviewers in de Nederlandse taal. *Eerste* generatie Marokkkanse respondenten werden ondervraagd door een Marokaanse interviewer in de Marokkaans-Arabishe taal. Van de tweede generatie Marokkaanse respondenten, werden 94 geïnterviewd in het Nederlands door een Nederlandse interviewer, 93 door een Marokkaanse interviewer in het Nederlands en 96 door een Marokkaanse interviewer in het Marokkaans-Arabisch. De tweede generatie Marokkaanse respondenten werden random toegewezen aan de drie hier genoemde interviewcondities.

3. RESULTATEN

3.1. *Kwaliteit van vragenlijsten*

3.1.1. *Resultaten Voeg-schaal analyses*

Met behulp van het Rasch model is de validiteit bepaald van een homogene suschaal van 8 items voor Nederlandse en Marokkaanse respondenten⁴. Onderzocht is of Nederlandse en Marokkaanse respondenten, die een gelijke mate van psychobiologische stress en fysiek ongemak ervaren, dezelfde kans hebben om de items bevestigend te beantwoorden. De kans om een item bevestigend te beantwoorden wordt, zoals eerder vermeld, uitgedrukt in de moeilijkheidsgraad van een item; hoe hoger de moeilijkheidsgraad van een VOEG-item, des te moeilijker is het om een item bevestigend te beantwoorden.

De resultaten in Tabel 1 laten zien dat 2 VOEG-items een significant verschillende moeilijkheidsgraad hebben in de Nederlandse en de eerste generatie Marokaanse respondentengroep (zie verder Van 't Land, 2000, hfd 2). Eerste generatie Marokkanen blijken systematisch vaker bevestigend te antwoorden op de het item “*Is uw maag nogal eens van streek?*” dan te verwachten was op basis van hun gezondheidstoestand (Tabel 1). Eerste generatie Marokkaanse zeggen dus gemakkelijker ‘ja’ op dit item dan vergelijkbare Nederlandse respondenten. Het item heeft dus in de Marokkaanse groep een lagere moeilijkheidsgraad ($p=-0.01$) dan in de Nederlandse groep ($p=1.50$). Als dit item gebruikt zou worden in minderhedensurveys dan zal geheel onterechte de gezondheidstoestand van eerste Marokkaanse respondenten als slechter ingeschat worden dan die van Nederlandse respondenten.

Deze analyses tonen aan dat deze subschaal minder geschikt is om de mate van psychobiologische stress en fysiek ongemak te meten onder eerste generatie Marokkaanse

⁴ VOEG-item 1 is niet opgenomen in de analyses aangezien in de Nederlandse respondengroep geen variantie geregistreerd werd op dit item.

respondenten. Voor tweede generatie Marokkaanse respondenten geldt dit overigens niet; in deze groep blijkt de schaal op gelijke wijze psychobiologische stress en fysiek ongemak te meten als onder vergelijkbare groepen Nederlandse respondenten.

Tabel 1. Vergelijking tussen de Item Moeilijkheidsgraad en Item fit informatie van eerste generatie Marokkaanse respondenten en vergelijkbare groepen Nederlandse respondenten.

VOEG-items	Item Moeilijkheidsgraad Oudere Nederlandse Respondenten (N=95)	Item Moeilijkheidsgraad Eerste generatie Marokkaanse respondenten (N=99)	Item Fit (χ^2)
1 'eetlust'			
2 'vieze of zoete smaak'	0.87	1.23	0.34
3 'rug'	-1.48	-0.77	3.70
4 'hoofdpijn'	-0.34	-0.88	3.60
5 'duizelig'	0.72	-0.01	3.70
6 'maag'	1.50	-0.01	9.90*
7 'slaperig'	-0.85	-0.36	1.40
8 'tintelingen in ledematen'	-0.42	0.81	10.31*

* p<0.05

3.1.2. Ontwikkelen van een nieuwe vragenlijst

Diepte-interviews met Nederlandse en Marokkaanse informanten laten belangrijke cross-culturele verschillen zien in aspecten die belangrijk gevonden worden op de algemene levensdomeinen.

De resultaten van de pilotstudie laten dit beeld eveneens zien. Nederlandse respondenten vinden vooral de 'Nederlandse' aspecten (de aspecten die genoemd zijn door 'Nederlandse' informanten) belangrijk bij het veroorzaken van gevoelens van (on)tevredenheid op verschillende levensdomeinen en, zoals te verwachten was, hangen de corresponderende vragen het sterkst samen met hun algemene gevoelens van (on)tevredenheid van ten aanzien van de huisarts. Marokkaanse respondenten blijken daarentegen zowel 'Marokkaanse' als 'Nederlandse' aspecten belangrijk te vinden bij het veroorzaken van gevoelens van (on)tevredenheid op verschillende levensdomeinen. Wel is het zo dat vragen die correspondeerden met 'Marokkaanse' aspecten in de Marokkaanse groep ook het sterkst samenhangen met de algemene gevoelens van (on)tevredenheid van Marokkaanse respondenten over de huisarts.

In Tabel 2 wordt een overzicht gegeven van aspecten die volgens de Nederlandse en de Marokkaanse informanten van invloed zijn op de algemene (on)tevredenheidsgevoelens van Nederlandse en Marokkaanse respondenten over de huisarts. Uit dit overzicht blijkt dat Marokkaanse informanten vooral tastbare en meetbare aspecten belangrijk vinden (zoals 'Mijn huisarts moet mij niet zonder medicijnen weg laten gaan') terwijl Nederlandse informanten vooral belang hechten aan minder tastbare aspecten (zoals 'Mijn huisarts moet mij veel aandacht geven').

Tabel 2. Overzicht van aspecten die Nederlandse en Marokkaanse informanten noemden tijdens de diepte-interviews.

‘Nederlandse’ aspecten	‘Marokkaanse’ aspecten
<ul style="list-style-type: none"> • ‘Ik moet met mijn huisarts over persoonlijke zaken kunnen spreken’ • ‘Mijn huisarts moet mij veel aandacht geven’ • ‘Mijn huisarts moet mij vragen wat volgens mij de oorzaak van mijn ziekte is’ • ‘Mijn huisarts moet mij vragen welke behandeling ik zelf het liefst wil’ • ‘Mijn huisarts moet niet zeggen dat het allemaal nog erger had kunnen zijn’ • ‘Mijn huisarts moet onderzoeken in hoeverre mijn klachten veroorzaakt zijn door stress’ 	<ul style="list-style-type: none"> • ‘Mijn huisarts moet veel medicijnen voorschrijven’ • ‘Mijn huisarts moet mij niet zonder medicijnen weg laten gaan’ • ‘Mijn huisarts moet meteen rontgenfoto’s laten maken’ • ‘Nadat ik mijn klachten verteld heb, moet mijn huisarts meteen een recept uitschrijven’ • ‘Mijn huisarts moet mij meteen doorverwijzen naar een specialist’ • ‘Mijn huisarts moet veel tijd voor mij uittrekken’

Bovengenoemde aspecten zijn vertaald in survey vragen en afgenomen in een groot-schalig onderzoek onder 578 mannelijke Marokkaanse en Nederlandse respondenten (zie ook paragraaf 2 van dit artikel).

Belang van Nederlandse en Marokkaanse aspecten in een vragenlijst

Na afname van survey vragen die gebaseerd zijn op Nederlandse en Marokkaanse aspecten, zijn de antwoorden van de Marokkaanse en Nederlandse respondenten met regressie analyse technieken nader geanalyseerd. Resultaten van Nederlandse respondenten laten zien dat vragen die gebaseerd zijn op ‘Nederlandse’ aspecten het sterkst samen (beta=0.59) hangen met de algemene gevoelens van (on)tevredenheid van Nederlandse respondenten over de huisarts (Tabel 3).

Tabel 3. Resultaten van een ‘stepwise’ regressie analyse op het levensdomein ‘huisarts’ voor Nederlandse en Marokkaanse respondenten.

Vragen gebaseerd op:	Nederlandse respondenten (N=196) Beta p	Stepwise Regressie Analyse	Marokkaanse respondenten (N=382) Beta p	Stepwise Regressie Analyse
‘Nederlandse’ aspecten	0.59 0.00	Step 1 (R ² =0.34)	0.36 0.00	Step 2 (R ² =0.54)
‘Marokkaanse’ aspecten	0.25 0.00	Step 2 (R ² =0.38)	0.69 0.00	Step 1 (R ² =0.48)
Adjusted R ²		Adjusted R ² =0.38		Adjusted R ² =0.54

Nederlandse respondenten, zowel jong als oud, vinden het vooral belangrijk ‘dat een huisarts vraagt wat de patient zelf denkt dat de oorzaak van zijn klachten is’ (beta=0.21) en ‘dat je vertrouwelijke zaken met een huisarts bespreken kunt’

(beta=0.17). In de Marokkaanse groep spelen deze aspecten geen rol van betekenis. In deze groep hangen juist vragen die betrekking hebben op de 'Marokkaanse' aspecten het sterkst samen met de algemene gevoelens van (on)tevredenheid over de huisarts (beta=0.69). Om tevreden te zijn over een huisarts, vinden Marokkaanse respondenten het vooral belangrijk 'dat een huisarts veel tijd uittrekt voor een bezoek' (beta=0.28) en 'dat een huisarts je niet zonder medicijnen laat weggaan' (beta=0.23).

Ook in de tweede generatie Marokkaanse respondentengroep speelt de Marokkaanse cultuur nog steeds een belangrijke rol. Tweede generatie Marokkaanse respondenten laten hun oordeel over de huisarts zelfs in sterke mate bepalen door aspecten die in de Marokkaanse cultuur belangrijk gevonden worden. Vooral survey vragen die betrekking hebben op de 'Marokkaanse' aspecten 'dat een huisarts veel tijd moet uittrekken voor een bezoek' (beta=0.31) en 'dat een huisarts je niet zonder medicijnen moet laten weggaan' (beta=0.25) hebben de grootste invloed op hun gevoelens van (on)tevredenheid over hun huisarts.

In de eerste generatie Marokkaanse respondentengroep lijken survey vragen die betrekking hebben op 'Nederlandse' aspecten de grootste invloed uit te oefenen. Analyses tonen echter aan dat dit effect uitsluitend veroorzaakt wordt door een survey vraag die betrekking heeft op een 'Nederlands' aspect, dat zegt 'dat een huisarts dient te onderzoeken of een klacht door stress veroorzaakt wordt'. Twee survey vragen die betrekking hebben op de 'Marokkaanse' aspecten 'dat een huisarts meteen moet doorverwijzen naar een specialist' (beta=0.36) en 'dat een huisarts veel tijd moet uittrekken voor een bezoek' (beta=0.27), hebben ook in deze groep significante invloed.

Tabel 4. Resultaten van een 'stepwise' regressie analyse op het domein 'huisarts' voor eerste en tweede generatie Marokkaanse respondenten.

Vragen gebaseerd op:	Tweede generatie Marokkaanse respondenten (N=283) Beta p	Stepwise Regressie Analyse	Eerste Generatie Marokkaanse respondenten (N=99) Beta p	Stepwise Regressie Analyse
'Nederlandse' aspecten	0.32 0.00	step 2 R ² =0.58	0.47 0.00	step 1 R ² =0.41
'Marokkaanse' aspecten	0.49 0.00	step 1 R ² =0.53	0.27 0.00	step 2 R ² =0.46
Adjusted R ²		Adjusted R ² =0.58		Adjusted R ² =0.46

3.2. Taal- en interviewer effecten

Taal en interviewereffecten zijn op twee manieren onderzocht. Allereerst zijn CBS survey vragen ingedeeld in een van de volgende categorieën: 1) ethnisch gerelateerd met ruimte voor subjectieve interpretatie, 2) ethnisch gerelateerd met geen ruimte voor subjectieve interpretatie of 3) 'neutraal': noch ethnisch gerelateerd noch ruimte voor subjectieve interpretatie (zie verder voor een uitgebreid overzicht Van 't Land, 2000, hfd 6).

Voor ieder van de categorieën is nagegaan of tweede generatie Marokkaanse respondenten andere antwoorden gaven aan een Nederlandse dan aan een Marokkaanse interviewer en of de taal, die tijdens het interview gesproken werd, een rol van betekenis speelde. Hieronder worden per categorie de antwoorden van tweede generatie Marokkaanse respondenten op een survey vraag besproken.

3.2.1. Etnisch gerelateerd onderwerp met ruimte voor subjectieve interpretatie

Verondersteld werd dat significante effecten vooral te verwachten zijn bij survey vragen die betrekking hebben op etnisch gerelateerde onderwerpen waar ruimte voor subjectieve interpretatie aanwezig is. Een voorbeeld hiervan is een survey vraag die de ervaren etnische identiteit van een respondent meet. Bij het definiëren van de etnische identiteit van een respondent, wordt in onderzoek meestal de definitie van het ministerie van Binnenlandse Zaken gebruikt. Deze definitie is gebaseerd op *herkomst*. Dat wil zeggen dat iemand als Marokkaan beschouwd wordt als de persoon in Marokko geboren is of een van zijn beide ouders. Inmiddels gaan in onderzoekskringen stemmen op dat een dergelijk onderscheid geen recht doet aan de identiteit die de respondent zelf ervaart. De respondent zou *zelf* moeten kunnen aangeven tot welke groep hij zich rekent. Het CBS heeft een vraag naar ervaren etnische identiteit opgenomen in het Leef Situatie Onderzoek (LSO).

Ook in mijn onderzoek heb ik deze survey vraag gesteld aan tweede generatie Marokkaanse respondenten. Het antwoord dat Marokkanen op deze vraag gaven, blijkt sterk afhankelijk te zijn van de etnische identiteit van de interviewer. Tweede generatie Marokkaanse respondenten zeggen vaker zich ‘Marokkaan’ te voelen als de interviewer een Marokkaan is. Tegen een Nederlandse interviewer daarentegen zeggen tweede generatie Marokkanen juist vaker zich ‘Nederlander’ of ‘Beide’ te voelen (Tabel 5). De resultaten laten significante interviewer effecten zien ($\chi^2=46.43;df=2; N=187;p<0.05$): 47.9% van de respondenten, geïnterviewd door een Nederlandse interviewer, zegt zich Marokkaan te voelen tegenover 90.3% van de respondenten, die geïnterviewd zijn door een Marokkaanse interviewer.

Tabel 5. “Kunt u mij zeggen of u zich meer een Marokkaan of meer een Nederlander voelt of Beide?”

	Nederlandse Interviewer/ Nederlandse Taal Conditie 1 (N=94)	Marokkaanse Interviewer/ Nederlandse Taal Conditie 2 (N=93)	Marokkaanse Interviewer/ Marokkaans-Arabische Taal Conditie 3 (N=96)	Totaal
‘Nederlander’	15.9%	6.5%	7.3%	9.9%
‘Marokkaan’	47.9%	90.3%	87.5%	75.3%
‘Beide’	36.2%	3.2%	5.2%	14.8%

Op basis van deze resultaten kan uitsluitend geconcludeerd worden dat bijna de helft van de respondenten (47.9%) tegen een Nederlandse interviewer zegt zich ‘Marokkaan’ te voelen. Is de interviewer een Marokkaan, dan moet een geheel andere conclusie getrokken worden; dan is juist de conclusie gerechtvaardigd dat tweede generatie

Marokkanen zich zeer sterk identificeren met hun etnische afkomst. Welke combinatie van taal en etniciteit van de interviewer het *juiste* antwoord oplevert, blijft onduidelijk en moet in vervolgonderzoek uitgezocht worden. Niet automatisch mag echter aangenomen worden dat een Marokkaanse interviewer per definitie “betere” antwoorden krijgt. Het zou immers kunnen zijn dat het voor een tweede generatie Marokkaan bedreigend of sociaal *onwenselijk* is om tegen een Marokkaanse interviewer te zeggen dat hij zich Nederlander voelt of dat hij over twee etniciteiten beschikt.

3.2.2. *Etnisch gerelateerde onderwerp zonder ruimte voor subjectieve interpretatie*

De vraag “Heeft u in Marokko op school gezeten?” is een voorbeeld van een vraag die weliswaar betrekking heeft op een etnisch gerelateerde onderwerp maar geen ruimte voor subjectieve interpretatie toelaat. Verondersteld werd dat er geen significante effecten aangetroffen worden bij dit type survey vragen. De resultaten laten inderdaad geen significante verschillen zien ($\chi^2(2;283)=2.83;p<0.05$). Onafhankelijk van de etniciteit van de interviewer of de taal van het interview, zegt het merendeel van de tweede generatie Marokkaanse respondenten niet in Marokko op school gezeten te hebben.

Tabel 6. “Heeft u in Marokko op school gezeten?”

	Nederlandse Interviewer/ Nederlandse Taal Conditie 1 (N=94)	Marokkaanse Interviewer/ Nederlandse Taal Conditie 2 (N=93)	Marokkaanse Interviewer/ Marokkaans-Arabische Taal Conditie 3 (N=96)	Totale percentage
“nee”	69.1%	78.5%	78.1%	75.3%
“ja”	30.2%	21.5%	21.9%	24.7%

3.2.3. *Geen etnisch gerelateerd onderwerp en geen ruimte voor subjectieve interpretatie*

Een survey vraag zoals “Wat is uw leeftijd?” heeft noch betrekking op een etnisch gerelateerd onderwerp noch laat het ruimte voor subjectieve interpretatie. Bij dit type vraag werden dan ook geen significante taal- en interviewer effecten voorspeld. Zoals verwacht, verschillen de antwoorden van de respondenten ook niet wezenlijk van elkaar. Resultaten van een univariate variantie-analyse laten geen hoofdeffect zien ($F=2.84; p >0.05$) en significante interviewer- en taaleffecten zijn afwezig.

Tabel 7. “Wat is uw leeftijd?”

Nederlandse Interviewer/ Nederlandse Taal Conditie 1 (N=94)	Marokkaanse Interviewer/ Nederlandse Taal Conditie 2 (N=93)	Marokkaanse Interviewer/ Marokkaans-Arabische Taal Conditie 3 (N=96)
20.6 (sd=2.9)	20.7 (sd=2.4)	19.9 (sd=2.4)

3.2.4. Sorteertaak domein Opvoeding

Niet alleen zijn taal- en intervieweffecten onderzocht door antwoorden van respondenten op survey-vragen te analyseren, ook zijn deze effecten onder de loep genomen door resultaten van een sorteertaak nader te analyseren. In dit artikel worden de resultaten van een sorteertaak op het gebied van opvoeding besproken (voor een meer uitgebreide bespreking van de resultaten wordt verwezen naar Van 't Land, 2000, hfd 6). Opvoeding is een onderwerp dat bij uitstek bepaald wordt door de normen en waarden die in een cultuur belangrijk gevonden worden. Wat goed of fout gedrag voor een kind is kan sterk verschillen tussen culturen.

In dit onderzoek is aangetoond dat hoe 'Nederlandser' de interviewsituatie is (taal van het interview en de etniciteit van de interviewer), des te meer belang tweede generatie Marokkaanse respondenten toekennen aan 'Nederlandse' opvoedingsaspecten. Ook blijkt dat hoe 'Marokkaanser' de interviewsituatie is (taal van het interview en de etniciteit van de interviewer), des te meer belang tweede generatie Marokkaanse respondenten hechten aan 'Marokkaanse' opvoedingsaspecten.

De resultaten van de sorteertaak laten zien dat zowel de taal van het interview als de etniciteit van de interviewer invloed uitoefenen op de resultaten van de sorteertaak. Een repeated measures analyse laat een significant *Aspect x Respondent group* effect zien ($F=26.58$, $p<0.05$). De drie respondentgroepen blijken significant te verschillen in het belang dat zij toekennen aan 'Nederlandse' en 'Marokkaanse' aspecten.

Tabel 8. Gemiddeld belang toegekend aan aspecten op het domein 'opvoeding', variërend van 0 ('onbelangrijk') tot 2 ('zeer belangrijk').

	Nederlandse Interviewer/ Nederlandse Taal Conditie 1 (N=94)	Marokkaanse Interviewer/ Nederlandse Taal Conditie 2 (N=93)	Marokkaanse Interviewer/ Marokkaans-Arabische Taal Conditie 3 (N=96)	Totaal Gemiddelde
'Nederlandse' aspecten	1.22 (sd=0.32)	0.94 (sd=0.40)	0.85 (sd=0.41)	1.00 (sd=0.41)
'Marokkaanse' aspecten	0.78 (sd=0.34)	0.90 (sd=0.38)	1.01 (sd=0.34)	0.90 (sd=0.36)
Totaal Gemiddelde	1.00 (sd=0.17)	0.92 (sd=0.25)	0.93 (sd=0.25)	0.95 (sd=0.23)
<i>Vershil</i>	<i>0.44</i>	<i>0.04</i>	<i>-0.16</i>	

4. CONCLUSIES EN AANBEVELINGEN

De onderzoeksvragen, die in dit artikel centraal staan, kunnen als volgt beantwoord worden:

Kwaliteit van vragenlijsten:

- 1) Zijn survey vragen die uitsluitend referen naar zaken die belangrijk zijn in de Nederlandse cultuur even belangrijk voor Marokkaanse als voor Nederlandse respondenten?

Dit blijkt niet het geval te zijn. Survey vragen die betrekking hebben op aspecten die belangrijk zijn in de Nederlandse cultuur, zijn niet even belangrijk voor Nederlandse en Marokkaanse respondenten. Voor het adequaat meten van (on)tevredenheids-gevoelens van Marokkaanse respondenten over hun huisarts, zijn vragen die refereren naar aspecten die belangrijk zijn in de Marokkaanse cultuur onmisbaar.

- 2) Neemt de geschiktheid van een vragenlijst toe voor Marokkaanse respondenten indien cultuurspecifieke survey vragen (vragen die referen naar zaken die belangrijk zijn in de Marokkaanse cultuur) aan een vragenlijst toegevoegd worden?

Deze vraag moet positief beantwoord worden. Algemene gevoelens van (on)tevredenheid over de huisarts worden minder adequaat gemeten onder Marokkaanse respondenten als een vragenlijst uitsluitend vragen bevat die betrekking hebben op aspecten die belangrijk zijn in de Nederlandse cultuur. Vragen die gebaseerd zijn op aspecten die belangrijk zijn in de Marokkaanse cultuur, doet de geschiktheid van een vragenlijst toenemen onder eerste en tweede generatie Marokkaanse respondenten. Als gevolg hiervan moeten resultaten van minderhedensurveys, waar uitsluitend gebruik gemaakt wordt van vragen die relevant zijn in de Nederlandse cultuur, met zorg geïnterpreteerd worden. Het verdient de aanbeveling om voorafgaand aan een onderzoek eerst de betekenis van een te onderzoeken domein vast te stellen onder de doelgroep van het onderzoek. Indien noodzakelijk, dienen extra vragen toegevoegd te worden, die corresponderen met aspecten die de doelgroep belangrijk vindt en die in de oorspronkelijke vragenlijst ontbreken.

Taal- en interviewer effecten:

- 3) Hebben de etnische afkomst van de interviewer en de taal die tijdens het interview gesproken wordt, invloed op de antwoorden die tweede generatie Marokkaanse respondenten geven op survey vragen?

Ook deze vraag moet positief beantwoord worden. Vooral bij vragen die betrekking hebben op etnisch gerelateerde onderwerpen die ruimte bieden voor subjectieve interpretatie, zijn significante interviewer- en taaleffecten gevonden. De etniciteit van de interviewer en de taal van het interview oefenen geen invloed uit bij vragen die *niet* gerelateerd zijn aan etnisch gerelateerde onderwerpen. Survey onderzoekers kunnen derhalve bij dit type vraag Nederlandse interviewers inzetten om tweede generatie Marokkaanse respondenten te interviewen. Indien vragen refereren aan etnisch gerelateerde onderwerpen, dan verdient het de aanbeveling om eerst te vast te stellen in hoeverre een vraag de ruimte biedt voor subjectieve interpretatie. Als dit niet het geval is, dan hoeven geen significante taal- en interviewer effecten verwacht te worden en kunnen Nederlandse interviewers ingezet. Indien er echter wel ruimte is voor subjectieve interpretatie, dan kunnen significante taal- en interviewer effecten verwacht worden. Bij vragen die noch betrekking hebben op etnisch gerelateerd onderwerp noch ruimte laten voor subjectieve interpretatie, hoeven wederom geen significante taal- en interview-

wer effecten verwacht te worden en kan volstaan worden met een Nederlands interview corps en Nederlandse vragenlijst.

Het blijft echter onduidelijk welke combinatie van taal en etniciteit van de interviewer het “juiste” antwoord oplevert. Het systematisch variëren van de etniciteit van de interviewer en taal kan in ieder geval onthullen of er sprake is van taal- en interviewereffecten. Meer onderzoek dient uitgevoerd te worden teneinde vast te stellen welke combinatie van taal en etniciteit van de interviewer de “beste” antwoorden oplevert.

• *Praktische aanbevelingen*

Tenslotte heeft het onderzoek een aantal praktische aanbevelingen opgeleverd voor onderzoekers die onderzoek doen onder onder minderheden. In mijn onderzoek heb ik bijvoorbeeld onderzocht welke onderzoeksvorm haalbaar was, gegeven de doelgroep waar ik me op richtte. Voorafgaand aan mijn onderzoek heb ik de toepasbaarheid van de volgende drie onderzoeksvormen⁵ onderzocht:

- a) schriftelijk
- b) telefonisch
- c) face-to-face

Bij de keuze voor een geschikte onderzoeksvorm, speelde een aantal overwegingen een rol:

- Is onderzoeksvorm praktisch haalbaar gegeven de doelgroep?
- Is de onderzoeksvorm bedreigend voor de doelgroep?
- Heeft de onderzoeksvorm aanluiting bij de survey praktijk?
- Levert de onderzoeksvorm data op die relatief eenvoudig te analyseren zijn?

a) *schriftelijke enquête*

Deze onderzoeksvorm heeft een aantal belangrijke voordelen. Het belangrijkste voordeel is het ontbreken van de noodzaak om interviewers in te zetten die eventueel de antwoorden systematisch kunnen beïnvloeden. Verder biedt deze onderzoeksvorm de mogelijkheid om gegevens anoniem te verzamelen en zijn de kosten voor het uitvoeren van het veldwerk relatief laag. De praktische toepasbaarheid van deze onderzoeksmethode is echter laag omdat het percentage analfabeten bij eerste generatie Marokkanen nog steeds relatief hoog is. Verder is het ontbreken van de mogelijkheid om respondenten mondelinge toelichting te geven op het doel van het onderzoek of op de betekenis van survey vragen voor dit onderzoek een belangrijk nadeel. Ook heeft de onderzoeker geen controle op het invullen van de vragenlijst.

b) *telefonische enquête*

Belangrijke voordelen van telefonisch enquêteren zijn de relatief lage kosten voor het uitvoeren van het veldwerk, de controle op de afname van de interviews en de relatief simpele invoer van de antwoorden van respondenten. Telefonisch is het echter moeilijk om eerste generatie Marokkaanse respondenten vertrouwen te geven en rustig uitleg te geven over het doel van het onderzoek, mede gelet op het feit dat telefonische inter-

⁵ Onbekendheid en onervarenheid van eerste generatie Marokkaanse respondenten met internet maakten dat deze onderzoeksvormen bij voorbaat als moeilijk toepasbaar beschouwd werd.

views maximaal twintig minuten mogen duren. De ervaring leert echter dat onbekendheid en onervarenheid met gestandaardiseerde interviews er voor zorgen dat interviews met eerste generatie Marokkaanse respondenten gemiddeld langer duren dan met Nederlandse respondenten. Telefonisch enquêteren is daarom voor dit onderzoek een minder geschikte onderzoeksvorm om in te zetten.

c) face-to-face enquête

Hoewel deze onderzoeksvorm het duurst is om uit te voeren, verdient deze vorm veruit de aanbeveling. De nadelen (hogere kosten voor het veldwerk, toename kans op systematische interviewer effecten en de afwezigheid van controle op afname interview) wegen uiteindelijk niet op tegen de voordelen van deze onderzoeksvorm. Het face-to-face interviewer biedt interviewers namelijk de mogelijkheid om toelichting te geven op het doel van het interview, eventuele ongerustheid bij de respondent weg te nemen en de duur van het interview desgewenst aan te passen aan het tempo van de respondent.

LITERATUUR

- Bronner, A.E. (1988). Surveying ethnic minorities. In: W.E. Saris & I.N. Gallhofer (Eds.), *Sociometric Research*. London: Macmillan.
- Choenni, C.E.S. (1984). 'Tien stellingen: minderhedenonderzoek ter discussie'. *Span'noe*, 26-27.
- Dirken, J.M. (1969). *Arbeid en stress. Het vaststellen van aanpassingsproblemen in werksituaties*. Groningen: JB Wolters.
- Heelsum, A.J van. (1997). *De etnisch culturele positie van de tweede generatie Surinamers*. Amsterdam: Het Spinhuis.
- Land, van 't H. (2000). *Similar Questions; Different Meanings. Differences in the meaning of constructs for Dutch and Moroccan respondents; effects of the ethnicity of the interviewer and language of the interview among first and second generation Moroccan respondents*. Leidschendam: KPN Research.
- Land, van 't H., Bosveld, W., Dijkstra, W. & Zouwen, J. van der (1994). Regarding the cultural background of a respondent: Effects on the content validity of questionnaires. In: *Proceedings of The International Conference on Survey Measurement and Process Quality*. Alexandria, VA: American Statistical Association.
- Land, van 't H., Bosveld, W., Dijkstra, W. & Zouwen, J. van der (1996). *Regarding the cultural background of a respondent: Item bias analysis using latent trait measurement theory*. Paper presented at the Fourth International ISA Conference on Social Science Methodology (Essex'96), Essex, april 1996.
- Liu, W.T. (1982). *Methodological problems in minority research*. s.l.: Pacific/Asian American Mental Health Research Centre.
- Meloen, J.D. & Veenman, J. (1990). *Het is maar de vraag...; Onderzoek naar responseffecten bij minderhedensurveys*. Lelystad: Koninklijke Vermande BV.
- Ministerie van Binnenlandse Zaken (1983). *Minderhedennota 1983*. Den Haag: Staatsuitgeverij.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Roelandt, Th. & Veenman, J. (1987). *Beter Meten 1. Aanbevelingen bij het rapportagesysteem Toegankelijkheid en Evenredigheid*. Rotterdam: ISEO.
- Veld, in 't R. & Steenbergen, J.E van. (1992). *Kleurrijk Utrecht nader bekeken. Nadere analyse van de gezondheidssituatie van allochtonen in de stad Utrecht, op basis van de gezondheidsspeiling 1986/87*. Utrecht: Gemeentelijke Geneeskundige en Gezondheidsdienst Utrecht.

6. *Zijn in marktonderzoek gevonden culturele verschillen echt?*

*Het bepalen van equivalentie in internationaal onderzoek middels detectie van bias**

H. VAN HERK

SAMENVATTING

Het is een basisvoorwaarde bij internationaal vergelijkend marktonderzoek dat de te vergelijken gegevens vergelijkbaar of equivalent zijn. In de literatuur wordt dit ook onderkend, hetgeen blijkt uit meerdere conceptuele artikelen die over dit onderwerp verschenen zijn. Echter, in het empirisch internationaal onderzoek wordt het vaststellen van equivalentie van gegevens doorgaans achterwege gelaten. Een verklaring hiervoor kan zijn dat de beschrijvingen van equivalentie in de literatuur veelal verwarrend en onduidelijk zijn. Bovendien is niet duidelijk hoe de diverse vormen van equivalentie vast te stellen zijn. In dit artikel wordt een procedure voorgesteld, die deze duidelijkheid poogt te geven. Bias wordt geïntroduceerd als een factor die de equivalentie van onderzoeksgegevens bedreigt. Er wordt aangegeven waar bias kan optreden in het onderzoeksproces en hoe bias geanalyseerd kan worden. Als er geen bias is, zijn gegevens equivalent en zijn cross-nationale vergelijkingen mogelijk. Tot slot van het artikel worden aanbevelingen gedaan ter verbetering van internationaal vergelijkend marktonderzoek.

1. INLEIDING

Bedrijven zoals Unilever, Philips, Ford en IKEA en vele, andere ondernemingen hebben een internationale strategie ontwikkeld en ook voor middelgrote ondernemingen wordt de internationale omgeving steeds belangrijker. Het toegenomen belang van internationalisering blijkt ook uit de toename van de internationale handel. In de afgelopen vijftig jaar is deze maar liefst gegroeid met een factor 17. Als gevolg hiervan is internationaal marktonderzoek steeds belangrijker geworden. Bij deze internationale onderzoeken worden steeds impliciet dan wel expliciet vergelijkingen tussen landen en culturen gemaakt.

Als vergelijkingen tussen landen gemaakt worden is “equivalentie” van marktinformatie essentieel. Equivalentie bestaat als informatie uit verschillende landen direct met

* Dit artikel is gebaseerd op het proefschrift “Equivalence in a Cross-National Context: Methodological and Empirical Issues in Marketing Research” (2000) van Hester van Herk; promotores: Y.H. Poortinga en Th. M.M. Verhallen.

elkaar vergeleken kan worden. Bijvoorbeeld, het aantal inwoners in Duitsland en Italië is direct met elkaar te vergelijken. Echter, het vaststellen van equivalentie is niet altijd voor de hand liggend; kennis over locale gewoonten kan noodzakelijk zijn om te weten of een begrip dezelfde betekenis heeft of niet. Bijvoorbeeld, oppervlakkig gezien lijkt het alsof 'frituren' in Frankrijk en Griekenland dezelfde betekenis heeft. Toch is dat niet zo. Het kookproces is totaal verschillend. In Griekenland wordt olijfolie in relatief kleine hoeveelheden gebruikt, terwijl in Frankrijk een relatief grote hoeveelheid plantaardige olie wordt gebruikt voor dezelfde toepassing. Dit heeft tot gevolg dat directe vergelijkingen tussen Frankrijk en Griekenland over attitudes ten aanzien van 'frituren' niet gemaakt kunnen worden. Men heeft het namelijk over een ander proces met andere uitkomsten. Zulke verschillen tussen landen kunnen in internationaal onderzoek leiden tot fouten, die in onderzoek in één land niet bestaan dan wel veel minder invloed hebben. Vergelijkbaarheid of equivalentie van (markt)informatie is van vitaal belang in internationaal onderzoek, omdat niet-equivalente gegevens kunnen leiden tot ambigue of zelfs foute conclusies.

Dit laatste is ook onderkend: de equivalentie van gegevens die in verschillende landen verzameld zijn wordt beschouwd als een basisvoorwaarde bij internationaal vergelijkend onderzoek (Sekaran, 1983; Hui en Triandis, 1985; Singh, 1995; Van de Vijver en Leung, 1997a,b; Craig en Douglas, 2000). Echter, ondanks het eraan toegekende belang wordt de equivalentie van gegevens meestal niet onderzocht (Aulakh en Kotabe, 1993) en de meeste studies noemen equivalentie zelfs helemaal niet. Eén van de redenen hiervoor kan zijn dat beschrijvingen van equivalentie en de gehanteerde begrippen in de literatuur veelal verwarrend en onduidelijk zijn. Bovendien geldt dat niet duidelijk wordt uit de literatuur hoe equivalentie is vast te stellen. Beide zaken hebben ertoe geleid dat het vaststellen van equivalentie doorgaans achterwege wordt gelaten zowel in academisch als in toegepast (markt)onderzoek.

Er lijkt behoefte te bestaan aan een procedure, die duidelijk aangeeft hoe equivalentie onderzocht kan worden. In dit artikel wordt hiervoor een aanzet gegeven. Bias wordt geïntroduceerd als een factor die de equivalentie van internationaal vergelijkend onderzoek bedreigt. Bias verwijst specifiek naar meetproblemen die kunnen ontstaan in het onderzoeksproces. Meetproblemen kunnen bijvoorbeeld optreden bij de vertaling van een vragenlijst, bij de sectie van respondenten en bij het interviewen. Het basisidee in dit artikel is dat wanneer de mate waarin bias de onderzoeksuitkomsten beïnvloedt bekend is, de mate waarin er equivalentie bestaat vast te stellen is. Namelijk, indien gegevens vrij zijn van bias dan zijn zij equivalent.

In de volgende paragrafen wordt besproken wat equivalentie is en hoe het in de huidige literatuur voorkomt. Vervolgens wordt aandacht besteed aan het detecteren en analyseren van bias, waarna wordt besproken hoe onderzoekers de equivalentie van hun gegevens kunnen vaststellen. Tot slot worden conclusies en aanbevelingen gegeven.

2. TYPEN STUDIES IN INTERNATIONAAL MARKTONDERZOEK

In het algemeen kunnen er in internationaal marktonderzoek drie typen studies worden onderscheiden. Ten eerste, studies die speciaal voor elk land apart worden ontworpen.

In zulke studies zullen landspecifieke zaken goed tot hun recht komen. Sommige onderzoekers stellen dat dit een zeer zuivere manier van internationaal onderzoek is (“emic” school; zie bijvoorbeeld Craig en Douglas, 2000: 156). Het landspecifieke heeft echter tot gevolg dat de vragenlijsten met grote waarschijnlijkheid zullen verschillen van elkaar. Het maken van vergelijkingen tussen landen zal dan moeizaam, zo niet onmogelijk zijn. In marktonderzoek zijn deze studies niet populair, omdat internationale marketeers altijd vergelijkingen tussen landen willen maken.

Ten tweede studies waarbij per land aanpassingen worden gemaakt aan de opzet van de studie en de vragenlijst. Zulke aanpassingen komen vaak voor bij internationaal onderzoek in de praktijk. Redenen voor aanpassing kunnen zitten in het niet beschikbaar zijn van dezelfde merken bij vragen naar merkbekendheid of in het gebruiken van ‘de eigen set attitudevragen’ in een land. Dit laatste gebeurt om vergelijkbaarheid met eerdere metingen in het specifieke land mogelijk te maken; echter, cross-nationale vergelijkingen worden hierdoor sterk bemoeilijkt. Veelal is bij deze studies vergelijking ook niet (meer) mogelijk. Een algemeen geldende regel is dat de vragen die veranderd zijn niet vergeleken kunnen worden over landen heen. Het betekent dat alleen het deel dat dezelfde vragen bevat (vertalingen uitgezonderd) vergeleken kan worden. Dus, in de situatie dat een studie aanpassingen op land niveau bevat zijn cross-nationale vergelijkingen gecompliceerd.

Het laatste type studie, het vergelijkende onderzoek, is speciaal ontworpen om vergelijkingen tussen landen te maken. Dit artikel zal zich alleen richten op deze vergelijkende onderzoeken. Op basis van deze studies kan vergeleken worden tussen landen, mits equivalentie vastgesteld is.

3. EQUIVALENTIE KADERS

3.1. Equivalentie in de marketing literatuur

Het kader voor equivalentie dat in marketing steeds weer aangehaald wordt is dat van Craig en Douglas (1983, 2000). Zij definiëren equivalentie als “Data that have, as far as possible, the same meaning or interpretation, and the same level of accuracy, precision of measurement, or reliability in all countries and cultures” (Craig en Douglas, 2000:141)¹. Hun equivalentiekader is breed gedefinieerd. Het omvat aspecten die met meten te maken hebben, maar ook het te bestuderen onderwerp, elementen uit het onderzoeksproces (bijv. vertalen), zaken uit de onderzoeksomgeving en de analyses. Dit kader geeft daarmee een verzameling weer van zaken waar rekening mee gehouden moet worden bij het doen van internationaal vergelijkend onderzoek. Craig en Douglas noemen drie hoofd vormen van equivalentie, construct equivalentie, meet equivalentie en steekproef equivalentie. Zij verwijzen hiermee naar de inhoud van een onderzoek, naar de wijze van dataverzameling en analyse en naar de wijze van steekproeftrekking en keuze van respondenten.

¹ Dit is dezelfde definitie als degene die Douglas en Craig in de eerste editie van hun boek “International Marketing Research” (1983) gebruikten.

Het kader voor equivalentie van Craig en Douglas is veelomvattend. Een belangrijk nadeel van hun kader is dat er niet duidelijk gemaakt wordt *hoe* equivalentie vast te stellen is. Zij stellen echter wel dat equivalentie moet bestaan voordat vergelijkingen gemaakt kunnen worden. De weg erheen is echter niet duidelijk. In de cross - culturele psychologie is door Van de Vijver en Leung (1997a,b) een kader gemaakt dat wel handvatten geeft ter bepaling van equivalentie. Dit kader wordt in de volgende paragraaf beschreven.

3.2. *Equivalentie in de cross-culturele psychologie*

In de cross-culturele psychologie is de laatste jaren veel aandacht geschonken aan het vaststellen van equivalentie. Bijvoorbeeld, Van de Vijver en Leung (1997) onderscheiden drie vormen van equivalentie: construct equivalentie, equivalentie van meeteenheid en scalaire equivalentie. Alle drie hebben zij betrekking op *metingen* in internationaal vergelijkend onderzoek. Equivalentie omvat in dit kader geen inhoudelijke aspecten van het onderzoek en geen zaken betreffende de steekproef. Dit betekent bijvoorbeeld dat zaken als vaststellen van de juiste productcategorie in alle landen bij Van de Vijver en Leung niet onder het begrip 'equivalentie' vallen.

De drie vormen van equivalentie die genoemd worden door Van de Vijver en Leung zijn aan elkaar gerelateerd. Equivalentie van meeteenheid kan alleen gelden als er construct equivalentie bestaat en scalaire equivalentie kan alleen bestaan als er construct equivalentie én equivalentie van meeteenheid bestaat. Zij beschrijven de equivalentie vormen als volgt:

- 1) *Construct equivalentie*. Constructequivalentie verwijst naar gelijkheid van psychometrische eigenschappen van gegevens uit verschillende landen. De meerderheid van de academische literatuur richt zich op het bepalen van construct equivalentie en in deze studies wordt nagegaan of er gelijke factor structuren bestaan in de verschillende landen. Een voorbeeld van een construct dat over landen heen equivalent is, is de CETSCALE (Shimp en Sharma, 1987). Deze schaal, die bestaat uit 17 items en etnocentrisme meet, bleek construct equivalent te zijn in de Verenigde Staten en Korea (Netemeyer, Durvasula en Lichtenstein, 1991).
- 2) *Equivalentie van meeteenheid*. Equivalentie van meeteenheid verwijst naar een situatie waar de meeteenheid gelijk is over landen heen, maar waar het nulpunt van de meetschaal verschilt. Een voorbeeld uit de fysica is het meten van de temperatuur in graden Celsius en Kelvin. Eén graad verschil is een zelfde verschil, maar het nulpunt is totaal verschillend. In het marktonderzoek bestaan geen meetinstrumenten zoals thermometers. Maar, ook bij meetinstrumenten in het marktonderzoek kunnen verschillen in niveaus van scores ontstaan, doordat er bias in de scores zit. Bijvoorbeeld, respondenten in het ene land kunnen veel meer de neiging hebben om yeasaying te vertonen dan respondenten in een ander land (zie bijvoorbeeld, Lee en Green 1991; Chen, Lee en Stevenson, 1995; Van Herk, 2000). Als er equivalentie van meeteenheid is dan kunnen scorepatronen, zoals twee metingen van dezelfde variabele in de tijd, tussen landen vergeleken worden. Er kan bijvoorbeeld geconcludeerd worden dat de koopintentie in beide landen met 0,5 punt is gestegen. Dat is een even grote stijging. Echter, scores op 1 variabele die in twee landen maar 1 maal gemeten is, kan alleen *binnen* de betreffende landen vergeleken worden. Een

vergelijking *tussen* landen op basis van scores mag alleen als het nulpunt van de schalen ook gelijk is in beide landen.

- 3) *Scalaire equivalentie*. Scalaire equivalentie bestaat als niet alleen de meeteenheid gelijk is over landen heen, maar ook het nulpunt. Als dit niveau bereikt is kunnen koopintentie scores direct vergeleken worden tussen respondenten uit verschillende landen. Het betekent dat een persoon met een score een '4' in het ene land dezelfde koopintentie heeft als een persoon met een score '4' in het ander land.

Van de Vijver en Leung richten zich in hun aanpak van equivalentie specifiek op de laatste fase in het onderzoeksproces, analyseren en interpreteren van gegevens. Equivalentie is voor hen een niveau dat bereikt kan worden en dat niveau bepaalt in hoeverre er vergelijkingen tussen landen te trekken zijn. Zij definiëren equivalentie als dat deel van het cross-culturele onderzoeksproces dat *achteraf* te bepalen is met behulp van statistische analyse procedures. In deze definitie worden zaken in de design fase van een studie, zoals allocatie van producten aan productcategorieën, vertaling van vragenlijsten en de wijze van steekproeftrekken beschouwd als mogelijke bronnen van bias.

Van de Vijver en Leung stellen, in tegenstelling tot anderen in de (marketing) literatuur, dat equivalentie niet a-priori vastgesteld kan worden. Volgens hen kan enig *bewijs* van equivalentie kan alleen worden verkregen door middel van statistische analyse procedures nadat alle gegevens verzameld zijn. Equivalentie kan beschouwd worden als een te bereiken eindniveau. Geen enkele vorm van equivalentie kan vastgesteld worden tijdens het onderzoeksproces. Wat wel vastgesteld kan worden tijdens het onderzoeksproces is *bias*. Bias kan optreden bij de definitie van het construct, bij het vertalen van de vragenlijst, tijdens het interview en bij het coderen van de data. Als er sprake is van bias wordt het te bereiken equivalentieniveau verlaagd; in de volgende paragraaf wordt dieper ingegaan op verschillende typen bias.

4. BIAS

Bias is een algemene term voor allerlei versturende factoren die de validiteit van vergelijkend onderzoek bedreigen. Bias kan resulteren in verschillen in testresultaten tussen twee groepen mensen, terwijl in werkelijkheid geen verschillen tussen deze twee groepen bestaan. Een bekend voorbeeld van bias heeft betrekking op intelligentietests. In de Verenigde Staten zijn de verschillen in test scores op intelligentietests tussen blanke en Afro-Amerikanen toegeschreven aan test bias en niet aan werkelijke verschillen in intelligentie tussen deze bevolkingsgroepen (Landers, 1986). Oorzaken van bias zitten veelal in vragen die voor een groep respondenten misplaatst of onduidelijk zijn. In de volgende paragraaf worden verschillende typen bias beschreven.

4.1. Typen bias

Er worden in de cross-cultureel psychologische literatuur drie typen bias onderscheiden (Van de Vijver en Poortinga, 1997). Dit zijn construct bias, methode bias en item bias. Deze drie vormen van bias representeren verschillende niveaus van bias: construct bias verwijst naar bias in het specifieke onderwerp of het domein dat bestudeerd wordt, methode bias verwijst naar bias in het meetinstrument of de meetechalen (bijvoorbeeld

interviewer bias of response stijlen) en item bias verwijst naar bias in bepaalde items in het instrument. Bias verlaagt het te bereiken niveau van equivalentie.

Construct bias komt zeer vaak voor als een construct wordt bestudeerd in verschillende landen. Vaak worden studies eerst gedaan in de Westerse Wereld en vervolgens worden deze gerepliceerd in andere Westerse landen of in niet-westerse landen; dit maakt het waarschijnlijk dat er construct bias optreedt. Een bekend voorbeeld dat laat zien dat de definitie van constructen uitgebreid moest worden zijn de dimensies van Hofstede (Hofstede, 1980). In de westerse wereld waren vier dimensies gevonden, maar bij het uitvoeren van een zelfde studie in Azië (Chinese Culture Connection, 1987) bleek dat er een vijfde dimensie toegevoegd moest worden, Lange Termijn Oriëntatie; een dimensie die niet gerelateerd was aan een van Hofstede's dimensies. Construct bias kan ook voorkomen doordat er een gebrek aan overlap is tussen gedragingen in verschillende culturen. Bijvoorbeeld, in marketing hoeft een zelfde product niet dezelfde functie te vervullen in verschillende landen en als gevolg hiervan zullen mensen anders reageren op vragen over dat product. Het gebruik van fietsen voor recreatie en voor woon - werk verkeer kan niet direct vergeleken worden en hetzelfde geldt voor het gebruik van boter voor bakken en voor smeren op brood (Van Herk, Verhallen en Barzilay, 1994). Dus, construct bias komt voor als het construct dat gemeten wordt niet hetzelfde is over landen heen en / of als gedragingen die bij het construct behoren niet overlappen.

Methode bias verwijst naar gevallen waar alle, of de meeste items op een zelfde manier worden beïnvloed door een factor die onafhankelijk is van het bestudeerde construct. Voorbeelden zijn response stijlen zoals yea-saying en extreem antwoordgedrag (bijv. Greenleaf, 1992; Van Herk, 2000; Baumgartner en Steenkamp, 2001). Er is sprake van een yea-saying response stijl als respondenten ongeacht het onderwerp positief reageren op vragen in een vragenlijst. Yea-saying leidt ertoe dat de gemiddelde scores van respondenten die deze response stijl vertonen systematisch hoger liggen dan scores van respondenten die deze response stijl niet vertonen. Methode bias kan ook ontstaan door de interviewer (interviewer - respondent interactie), de onderzoeksmethode (telefonisch, schriftelijk of persoonlijk interviewen), of achtergrondkenmerken van respondenten in de steekproef zoals leeftijd of sociale klasse. Methode bias is doorgaans moeilijk op te sporen, omdat alle items in een vragenlijst gelijkmatig door deze vorm van bias worden beïnvloed. Als gevolg hiervan kunnen methode bias en 'echte' verschillen in bijvoorbeeld attitudes niet gemakkelijk gescheiden worden.

Zowel construct als methode bias hebben invloed op het totale onderzoeksinstrument. Item bias daarentegen betreft slechts één of enkele items in een multi-item schaal. In academisch marktonderzoek worden veel multi-item schalen gebruikt (zie Bearden en Netemeyer, 1999); voorbeelden zijn de CETSCALE (Shimp en Sharma, 1987) en de EAP schaal (Baumgartner en Steenkamp, 1996). Item bias treedt op wanneer in een land één of een klein aantal items in een multi-item schaal afwijken van het verwachte antwoordpatroon op deze multi-item schaal. Een voorbeeld kan dit illustreren. Stel, we gebruiken een schaal die 'gezondheidsbewustzijn' meet in twee landen en daarin zit de vraag 'ik bezoek minstens eenmaal per week een fitnessclub'. Stel bovendien dat de respondenten uit de twee groepen die in het onderzoek betrokken zijn in werkelijkheid een net zo groot gezondheid bewustzijn hebben. Als nu in een van de twee landen een beperkt aantal fitnessclubs is, dan is het aannemelijk dat men in dat land veel vaker

‘nee’ antwoordt in plaats van ‘ja’, gegeven een bepaald niveau van gezondheidsbewustzijn. In het andere land zal men ‘ja’ antwoorden zoals verwacht. In deze situatie is er sprake van item bias voor het item betreffende bezoek aan een fitnessclub. De oorzaken van item bias kunnen ook nog zitten in een inadequate formulering van de vraag (gebruik van moeilijke woorden bijvoorbeeld) of in een slechte vertaling.

Het vóórkomen van bias is een belangrijke bedreiging voor de validiteit van cross-nationale en cross-culturele vergelijkingen in marktonderzoek. Als er bias is zijn data niet equivalent en kunnen er geen valide cross-culturele vergelijkingen gemaakt worden.

4.2. Bias en equivalentie

Construct equivalentie bestaat als er geen construct bias in de data aanwezig is. Pas wanneer construct equivalentie is vastgesteld mag er geconcludeerd worden dat een construct, bijvoorbeeld ‘gezondheidsbewustzijn’, hetzelfde is in de bestudeerde landen. Numerieke vergelijkingen zijn niet toegestaan bij het bestaan van alleen construct equivalentie; numerieke vergelijkingen mogen pas gemaakt worden als er minimaal equivalentie van meeteenheid is. In Tabel 1 kan afgelezen worden welk equivalentieniveau te bereiken is. Bijvoorbeeld als er construct bias is dan is er geen construct equivalentie te bereiken; de “Nee” in de tabel verwijst hiernaar. Als er uniforme methode bias is dan zijn zowel construct equivalentie als equivalentie van meeteenheid te bereiken (“Ja” in de tabel).

Tabel 1. Welk equivalentieniveau is te bereiken? (naar Van de Vijver en Leung, 1997a).

Aanwezige bias	Equivalentie niveau		
	Construct	Meeteenheid	Scalaire
Construct bias	Nee	Nee	Nee
Methode bias			
Uniform	Ja	Ja	Nee
Niet-uniform	Ja	Nee	Nee
Item bias			
Uniform	Ja	Ja	Nee
Niet-uniform	Ja	Nee	Nee

Bij methode en item bias kan het onderscheid gemaakt worden tussen uniforme en niet-uniforme bias (zie ook Van de Vijver en Leung, 1997a). Uniforme bias refereert aan bias die hetzelfde is voor alle scoreniveaus. Bijvoorbeeld, een weegschaal geeft consistent 1 kilo te veel aan. Niet-uniforme bias refereert aan invloed die niet gelijk is voor alle scoreniveaus. Stel, 10% te veel bij wegen. Dit betekent dat bij iets van 1 kilogram 1,1 wordt aangegeven, dus 0,1 kg te veel. Bij 10 kilo is dit verschil opgelopen tot 1 kg. Dus vergelijkingen worden bij hogere gewichten meer verstoord. Niet-uniforme bias verstoort metingen in ernstige mate en leidt met grote waarschijnlijkheid tot foutieve cross-nationale vergelijkingen. In het vervolg zal er bij het noemen van methode bias dan ook uniforme bias bedoeld worden.

Als er geen uniforme methode bias is en geen construct bias dan mag er geconcludeerd worden dat het te bestuderen onderwerp construct equivalent is en dat er equivalentie van meeteenheid bestaat. Dit betekent dat verschillen in scores vergeleken mogen worden, maar absolute waarden niet. Bijvoorbeeld er kan geconcludeerd worden dat er op het construct 'gezondheidsbewustzijn' een groter verschil zit tussen jongeren en ouderen in land A dan tussen jongeren en ouderen in land B. Echter, om te concluderen dat men in Land A een hoger niveau van 'gezondheidsbewustzijn' heeft dan in land B zouden de scores ook scalaire equivalentie moeten hebben. Scalaire equivalentie is echter niet vast te stellen door alleen analyse van bias; voor het vaststellen van scalaire equivalentie is vaststellen van externe validiteit noodzakelijk.

4.3. Bias en het onderzoeksproces

Om equivalente informatie te verkrijgen in de laatste fase van het marktonderzoekproces (zie Tabel 2) moet er aandacht besteed worden aan bronnen van bias in alle voorgaande fasen. Namelijk, bias kan ontstaan in elke fase van het marktonderzoekproces. Bronnen van bias kunnen liggen in de operationalisatie van het construct, de vragenlijst, de vertaling, de steekproeftrekking, het veldwerk en in het coderen van de data. Pas achteraf, in de data analyse fase kan men bepalen in hoeverre er afwezigheid van bias ofwel equivalentie is.

In internationaal marktonderzoek is bias een gegeven. In het gehele onderzoeksproces kan bias optreden. Bias kan al in de allereerste fase in het onderzoek geïntroduceerd worden, namelijk bij de probleemformulering. In latere fasen in het marktonderzoek kan eveneens bias optreden. Bias is cumulatief, hetgeen betekent dat bias die in de eerste fase optreedt in latere fasen niet tenietgedaan wordt.

Als een onderzoeker een onderzoek dat in het eigen land gedaan is rechtstreeks gaat omzetten naar een ander land kan (met grote waarschijnlijkheid) bias in het onderzoek sluipen. Twee voorbeelden kunnen dit illustreren. (1) Sommige statements in een multi-item schaal kunnen minder van toepassing zijn. Een statement als 'American products, first, last and foremost' (uit de CETSCALE, Shimp en Sharma, 1987) is ook wanneer deze goed vertaald wordt in Nederland minder van toepassing dan in de USA. Het kopen van een echt Nederlandse auto is onmogelijk. (2) Bias sluipt ook in een onderzoek als gedragingen in een land gebruikelijk zijn en in een ander land niet. In Nederland is fietsen naar school door kinderen gebruikelijk, terwijl kinderen in veel andere landen gebruik maken van een schoolbus. Zo'n schoolbus bestaat in Nederland niet.

Het ontstaan van bias is niet beperkt tot een of enkele fasen in het onderzoeksproces. Het kan in elke fase voorkomen. Een overzicht van de zeven fasen in het onderzoeksproces en bronnen van bias die hierbij kunnen optreden kan dit verduidelijken (zie Tabel 2).

4.3.1. Fase I: Probleemformulering

In fase I van het onderzoeksproces wordt het probleem geformuleerd en worden de doelstellingen van het onderzoek gedefinieerd. In een cross-nationale studie is het gebruikelijk om na te gaan of het onderwerp relevant is in alle te bestuderen landen. Bijvoorbeeld, het te bestuderen onderwerp wordt gedefinieerd op basis van desk research en discussies met collega onderzoekers. Vervolgens wordt het gedefinieerd in

Tabel 2. Het onderzoeksproces, equivalentie en bias.

	Fasen in het marktonderzoek proces	Bron van bias	Onderwerp	Prevalerende typen bias
I	Probleem formulering	Definitie van concepten Bepalen van de juiste product-categorie Bepalen functie van producten	Doel van de studie	construct
II	Research design	Operationalisatie Instrument design Vertaling	Type studie Type vragen Item selectie Type response formaat Vragenlijst design	construct methode item
III	Methode van data verzameling	Methode	Persoonlijk, post, telefonisch, e-mail	methode
IV	Steekproef trekking	Sampling	Doelgroep Sampling frame	construct methode item
V	Data verzameling	Fieldwork	Procedures Interviewer selectie Time frame	methode
VI	Coderen	Editen Coderen Calibratie	Data editen Data coderen	item
VII	Analysen en interpreteren van de data		Statistische procedures	

de internationale context. Het eerst definiëren in alle landen apart en vervolgens in een internationale context is belangrijk, omdat er bij definiëren in 1 land en vervolgens replicatie in andere landen het risico bestaat dat er bias in het onderzoek voorkomt. Dit kan komen door de impliciete aanname van een onderzoeker dat begrippen en/of gedragingen in het andere land gelijk zullen zijn aan die in het eigen land. Bijvoorbeeld men kan er impliciet van uit gaan dat producten tot een zelfde productcategorie horen of dat producten een zelfde functie vervullen in het andere land.

Essentieel is dat in fase I van het onderzoek de equivalentie van het onderwerp nog steeds aangenomen wordt. Toetsing of het inderdaad equivalent is kan pas na afloop van de studie. Wanneer in fase I blijkt dat het onderwerp niet vergelijkbaar is over landen heen dan moet er geconcludeerd worden dat cross-nationale vergelijkingen niet gemaakt kunnen worden.

4.3.2. Fase II: designfase

In fase II, de designfase, worden beslissingen genomen over het type studie, de operationalisatie van constructen, item selectie en het responseformaat. Als een construct niet op dezelfde wijze geoperationaliseerd kan worden in de te bestuderen landen, dan ontstaat in deze fase construct bias. Het is bijvoorbeeld mogelijk dat een construct als 'agressie' bestaat in alle landen, maar dat in het ene land verbale agressie veel vaker voorkomt en in het andere land fysieke agressie. Dit betekent dat de vragen die gesteld moeten worden om 'agressie' te meten niet hetzelfde kunnen zijn. Bij een zelfde vraagstelling zal er dan construct bias ontstaan².

Methode bias wordt geïntroduceerd in deze fase als het responseformaat om de vragen te beantwoorden bekender is bij de respondenten in een van de landen. Dit laatste is niet ongebruikelijk. Bijvoorbeeld, in de Verenigde Staten wordt een 20-puntsschaal vaak gebruikt (Kotabe en Helsen, 1998), terwijl in Frankrijk de voorkeur wordt gegeven aan 5-punts en 7-puntsschalen. Als er een 20-puntsschaal wordt voorgelegd aan de Franse respondenten zullen zij geneigd zijn om de extremen van de schaal te vermijden; dit leidt dan tot hogere scores voor de respondenten uit de Verenigde Staten, ongeacht de werkelijke intensiteit van hun mening.

Item bias wordt geïntroduceerd in fase II als een item om een construct te meten geschikt is in één land, maar totaal ongeschikt in een ander land. Hetzelfde treedt op als de vertaling van een of meer items onjuist of onmogelijk is, omdat er voor bepaalde woorden in de andere taal geen woorden met een zelfde betekenis bestaan. Overigens kunnen veel problemen rond vertalen al voorkomen worden als de richtlijnen voor vertalen (Van de Vijver en Hambleton, 1996) opgevolgd worden.

4.3.3. Fase III: Methode van dataverzameling

In fase III, bepaling van de methode van dataverzameling, kan methode bias in de studie geïntroduceerd worden als respondenten minder bekend zijn met een bepaalde methode van dataverzameling. Bijvoorbeeld in Westerse landen is gecomputeriseerd interviewen met CAPI en CATI methoden gebruikelijk, terwijl het compleet onbekend is in andere delen van de wereld. Een lagere bekendheid met een onderzoeksmethode beïnvloedt de resultaten. Een voorbeeld is te vinden in een artikel van Serpell (1979). Serpell liet kinderen uit de UK en uit Zambia patronen namaken met twee methoden, respectievelijk pen en papier en ijzerdraadjes. Het bleek dat de kinderen uit Zambia het hoogst scoorden op de methode met de ijzerdraadjes en de engelse kinderen het hoogst scoorden op de methode met pen en papier. De kinderen scoorden het hoogst op de methode waarmee zij het meest vertrouwd waren.

4.3.4. Fase IV: Steekproeftrekking

In fase IV worden de steekproefsamenstelling en het steekproefkader bepaald. Het kiezen van steekproeven die gelijk zijn met betrekking tot demografische kenmerken kan helpen om bias te reduceren. Bij representatieve steekproeven in zeer verschillende lan-

² Dit betekent dat het construct 'agressie' niet hetzelfde is in de verschillende landen. Als de te stellen vragen anders zijn dan is het construct niet gelijk. Alleen als dezelfde vragen gesteld kunnen worden om een construct te meten kan er na analyse van de resultaten construct equivalentie vastgesteld worden.

den zoals bijvoorbeeld India en Duitsland kan een gevonden verschil in attitude vele alternatieve verklaringen hebben. Het verschil kan bijvoorbeeld toegeschreven worden aan een verschil in cultuur (wat wellicht gewenst was), maar net zo goed aan een verschil in welvaart. Door nu in beide landen respondenten te kiezen met een zo gelijk mogelijke achtergrond wordt het aantal alternatieve verklaringen verminderd. Om deze reden worden in academisch onderzoek vaak studenten als respondent ingezet. Dit, omdat studenten doorgaans gelijke demografische kenmerken bezitten. Echter, ook dan is bias niet uitgesloten: studenten zijn zeker niet representatief voor hun land van herkomst. In sommige landen studeren mensen uit alle sociale klassen, terwijl in andere landen alleen mensen uit de hogere sociale klasse studeren. Een verschillende sociale achtergrond beïnvloedt attitudes en kan derhalve construct en item bias introduceren.

4.3.5. Fase V en VI: Dataverzameling en coderen

In fase V, de dataverzameling fase, kan methode bias worden geïntroduceerd doordat de gevolgde procedures, bijvoorbeeld de interviewer instructies, verschillen. Methode bias kan ontstaan tijdens het interview als respondenten meer geneigd zijn te praten over gevoelige onderwerpen bij bepaalde speciale interviewers. Bijvoorbeeld, vrouwen zouden meer geneigd kunnen zijn te vertellen over geweld dat hen is aangedaan tegen vrouwelijke dan tegen mannelijke interviewers. Ook kan de interactie tussen respondent en interviewer bias veroorzaken. Bijvoorbeeld, in bepaalde landen bestaat 'courtesy bias'. Dit type bias wordt veroorzaakt door het feit dat men in die culturen van mensen verwacht dat zij beleefd zijn tegen interviewers. Hun antwoorden geven dan derhalve geen beeld van hun ware meningen en opinies (Kumar, 2000).

In fase VI, het coderen kan eveneens bias optreden. De kans hierop is het grootst als het coderen niet op een centrale locatie uitgevoerd wordt. Bijvoorbeeld, als zonder overleg antwoorden in categorieën worden gecodeerd in Italië en in Nederland is het niet ondenkbaar dat de codeurs dezelfde antwoorden aan twee verschillende categorieën toedelen.

4.3.6. Fase VII: Analyse fase

Uit het voorgaande blijkt dat bronnen van bias in elke fase van het onderzoeksproces kunnen ontstaan. Zij zijn bovendien cumulatief. Bijvoorbeeld, bij gebruik van een computer gestuurde vragenlijst (bijv. met de CAPI methode) en het gebruik van een 5-punts rating schaal in één studie kan één ervan, of beide tezamen, de methode bias verhogen. Er geldt bovendien, dat bias die in fase I niet onderkend is, niet in een latere fase gecorrigeerd kan worden. Als een onderwerp onvergelykbaar is kunnen zaken als bijvoorbeeld een correcte dataverzameling en een perfecte vertaling dat niet meer corrigeren. Equivalentie bestaat dan niet en vergelijkingen tussen landen zijn dan met grote waarschijnlijkheid foutief.

Pas in de laatste fase (VII), de analysefase is het mogelijk de afwezigheid van bias te bepalen met behulp van statistische analyse procedures.

4.4. Detecteren van bias

Om bias in de data te kunnen opsporen is nodig dat al in de vragenlijst en tijdens de data verzameling rekening is gehouden met deze bronnen van bias, bijvoorbeeld door

extra vragen op te nemen in een vragenlijst. Vervolgens kan er met behulp van statistische procedures gekeken worden of er sprake is van bias.

4.4.1. Opnemen van extra vragen

De basis om later de aanwezigheid van bias te kunnen detecteren met behulp van statistische technieken ligt voor een groot deel in de fase van het researchdesign (fase II). In deze fase kan een onderzoeker besluiten extra vragen op te nemen die latere verschillen in uitkomsten tussen de onderzochte landen kunnen verklaren. Bijvoorbeeld, er kunnen gegevens opgenomen worden over de interviewer, of over de mate van bekendheid met het onderwerp van studie en dergelijke. Bias die ontstaan is door deze factoren kan dan achteraf worden toegewezen aan deze factoren. Bijvoorbeeld, een hoge score kan verklaard worden door grote bekendheid van de respondenten met het onderwerp.

Het opnemen van extra vragen in de dataset geldt voor analyse van zowel construct als methode als item bias. Construct bias kan veroorzaakt worden doordat gedragingen alleen voorkomen in bepaalde landen. Deze vorm van construct bias kan alleen onderzocht worden door (1) vooraf kwalitatief onderzoek te doen en vervolgens (2) door in het onderzoek te vragen naar bijvoorbeeld frequentie van voorkomen van dat gedrag. Een andere vorm van construct bias kan ontstaan, omdat het construct niet voldoende wordt afgedekt in de vragenlijst. Een breed veelomvattend construct wordt in de vragenlijst gerepresenteerd door een klein aantal vragen of zelfs maar een vraag. Bijvoorbeeld, als een construct zoals 'innovatief zijn' gemeten wordt met één vraag is het risico groot dat juist die vraag in een ander land anders wordt geïnterpreteerd. Bij het gebruiken van circa vijf vragen is de kans dat al deze vragen anders worden geïnterpreteerd veel kleiner. Bovendien geldt dat construct bias opgespoord kan worden met statistische procedures als meerdere items tezamen dat ene construct meten. Bij meten met behulp van één item is dit niet mogelijk.

Methode bias kan ontstaan in het researchdesign, omdat bepaalde respondenten minder bekend zijn met de methode. Om dit later te kunnen controleren is het opnemen van een vraag die informeert naar de bekendheid met de methode of het van tevoren laten oefenen met de methode essentieel. Bij de dataverzameling kan methode bias ontstaan door verschillende fysieke omstandigheden tijdens de interviews of door andere interviewers. Het opnemen van een variabele in de dataset, die aangeeft wat de omstandigheden waren (bijvoorbeeld lawaai of een lage temperatuur) en wie de interviewer was biedt mogelijkheden om achteraf te analyseren of deze variabelen verschillen tussen landen kunnen verklaren.

Item bias verwijst naar artefacten op het niveau van individuele vragen. Item bias is vooral goed op te sporen als er multi-item schalen worden gebruikt³. Bij het meten van zaken met één vraag is het scheiden van bias en een werkelijk cross-nationaal verschil moeilijk zo niet onmogelijk. Een oplossing kan nog gevonden worden in fysieke meting van het fenomeen waar mogelijk. Doch, bij bijvoorbeeld 'veel' olie gebruiken kan de definitie van 'veel' duidelijk verschillen per land.

³ Het Handbook of Marketing Scales (Bearden en Netemeyer, 1999) bevat vele tientallen multi-item schalen die uiteenlopende onderwerpen meten zoals bijvoorbeeld over innovativiteit, koop-involvement en exploratief koopgedrag

Concluderend, er geldt dat het belangrijk is om zaken die bias kunnen veroorzaken te benoemen en deze mee te nemen als vragen in de vragenlijst en/of als extra variabelen toe te voegen aan de dataset. Verder is het meten van constructen met meerdere variabelen een noodzakelijke voorwaarde om construct en item bias in vragenlijsten te kunnen onderzoeken.

4.4.2. Statistische procedures

Pas nadat alle gegevens verzameld zijn is het mogelijk met behulp van statistische analyses na te gaan of er bias is. De belangrijkste technieken om de verschillende vormen van bias op te sporen staan weergegeven in Tabel 3. Elk van deze technieken zal in het vervolg van deze paragraaf kort worden toegelicht voor nadere uitleg kan men de literatuur raadplegen.

Tabel 3. Toepassingen van technieken in internationaal onderzoek.

Bias	Statistische technieken	Voorbeeld van gebruik in cross-culturele context
Construct	Principale Componenten Analyse + Procrustes rotatie	Katigbak, Church en Akamine, 1996; Van Herk, 2000
	Confirmatory Factor Analysis	Steenkamp en Baumgartner, 1998;
	Multi-Dimensionale Schaaltechnieken	Schwartz, 1992
Methode	Variantie Analyse	Dawar en Parker, 1994
	Covariantie Analyse	—
	Hierarchical Linear Modeling	Ter Hofstede, Steenkamp en Wedel, 1999
Item	Item Response Theorie	Ryan, Horvath, Ployhart, Schmitt en Slade, 2000
	Variantie Analyse	—

Voor het controleren van afwezigheid van construct bias zijn diverse methoden beschikbaar. Principale componenten analyse gevolgd door Procrustes rotatie van de oplossingen en vervolgens het berekenen van een coëfficiënt van overeenstemming (bijv. Tucker's Phi) is een vaak gebruikte methode voor het bepalen van overeenkomsten tussen meer-dimensionale PCA of factorstructuren. Confirmatorische Factor Analyse (bijv. Steenkamp en Baumgartner, 1998) is vooral geschikt voor het bepalen van overeenkomsten tussen eendimensionale multi-item schalen in meerdere landen. Bij variabelen met een ordinaal meetniveau kunnen Multi-Dimensionale Schaaltechnieken (MDS; zie bijv. Van der Kloot, 1997) ook gebruikt worden voor het bepalen van de gelijkheid van configuraties over landen heen; de overeenkomst van MDS configuraties kan, net als bij PCA, bepaald worden met behulp van Procrustes rotatie gevolgd door het berekenen van een coëfficiënt van overeenstemming.

Voor het analyseren van methode bias worden technieken gebruikt die zich richten op het niveau (de hoogte) van de scores in de diverse landen. Met behulp van variantie analyse, covariantie analyse of met een t-test kan nagegaan worden of er verschillen bestaan op variabelen als interviewomstandigheden, gebruikte interviewprocedure en

dergelijke. Hierarchical Linear Modeling (bijv. Bryk en Raudenbush, 1992) is een meer recent ontwikkelde techniek die speciaal ontwikkeld is voor data die op meerdere niveaus verzameld zijn, bijvoorbeeld op het niveau van (1) land, en (2) individu. Gevonden effecten zijn hierbij te splitsen in een landeffect en een effect op individueel niveau. Bijvoorbeeld 'Individualisme' (Hofstede, 1980) is een effect op landniveau, en leeftijd een effect op individueel niveau. Een voorwaarde voor toepassing van HLM is wel dat er op het hoogste niveau (land) minimaal 10 en liever meer observaties zijn. Een artikel waarin de techniek wordt toegepast om response stijlen te vinden in een internationale context is van Baumgartner en Steenkamp (2001).

Voor het analyseren van item bias bij dichotome variabelen kan de Mantel-Haenszel procedure (bijv. Holland en Thayer, 1988) of Item Response Theorie (IRT) gebruikt worden. Ook bij gebruik van polytome rating schalen kunnen IRT- modellen gebruikt worden (zie bijv. Van der Linden en Hambleton, 1997). IRT modellen worden nauwelijks gebruikt in internationaal onderzoek, een uitzondering is het recente artikel van Ryan en anderen (2000). Een wat eenvoudiger methode voor de analyse van item bias bij data met interval of ratio meetniveau is variantie analyse. Bij het gebruik van variantie analyse wordt dan eerst voor elke respondent, ongeacht land van herkomst, zijn/haar totaal score op de vragen (die één construct meten) bepaald. Vervolgens worden de respondenten ingedeeld in groepen op basis van deze totaalscore. In de variantie analyse worden deze groepen en het land de onafhankelijke variabelen en de afzonderlijke vragen steeds de afhankelijke. Een vraag is dan niet gebiast als zowel het hoofdeffect van land als de interactie tussen land en scoregroep niet significant is (zie ook Van de Vijver en Leung, 1997).

5. CONCLUSIE EN AANBEVELINGEN

Het gebruik van de term equivalentie voor zaken die betrekking hebben op diverse zaken die spelen bij het doen van internationaal marktonderzoek heeft veel verwarring geschapt. Deze verwarring wordt voorkomen door de term equivalentie alleen te gebruiken om aan te geven in hoeverre vergelijking van onderzoeksresultaten uit verschillende landen mogelijk is. In dit artikel wordt dit gedaan door te stellen dat equivalentie alleen achteraf middels data analyse vast te stellen is. Alle mogelijke verstoringen in internationaal vergelijkend onderzoek (zoals bijvoorbeeld fouten bij vertalingen) worden gevat onder de term 'bias'. Afwezigheid van bias duidt op equivalentie. Het verbinden van bias aan de fasen in het onderzoeksproces biedt onderzoekers een procedure, die aangeeft hoe equivalentie onderzocht kan worden en hoe het vast te stellen is. Reductie van bias verhoogt het niveau van equivalentie dat bereikt kan worden.

Als er geen bias in de scores is zijn data equivalent en mag er vergeleken worden over landen heen. De aanwezigheid van bias kan vastgesteld worden door opnemen van, onder andere, extra vragen in een vragenlijst en statistische analyses. Op basis van de analyses is het bereikte niveau van equivalentie vast te stellen: bijvoorbeeld, als er geen construct bias is, is er construct equivalentie.

Om enerzijds bias te kunnen reduceren en anderzijds te kunnen corrigeren is een drietal aanbevelingen te doen. Ten eerste wordt het gebruik van multi-item schalen aan-

bevolen om construct bias te kunnen beperken. Bij multi-item schalen is reductie van bias een reële optie⁴. Bij meting met slechts 1 of 2 variabelen rest na constatering van bias geen andere mogelijkheid dan afzien van vergelijking tussen de betreffende landen. Ten tweede, door meer gebruik van kwalitatief onderzoek kan bias in de design fase van het onderzoek gereduceerd worden. Informatie uit kwalitatief onderzoek kan dienen als input voor kwantitatief onderzoek. Bijvoorbeeld, in kwalitatief onderzoek kan duidelijk worden dat de gebruikssituaties van een product per land verschillen. In de designfase bij een kwantitatief vervolgonderzoek kunnen dan extra vragen worden opgenomen over de gebruikssituaties. Tenslotte, wordt het verzamelen van gegevens over het onderzoeksproces (bijv, wie interviewt) en over de omgeving (bijv, gemiddelde gebruiksfrequentie van het te onderzoeken product) aanbevolen. Het verzamelen van extra gegevens kan helpen om gevonden cross-culturele verschillen te verklaren. Het kan dan blijken dat een gevonden verschil te verklaren is uit verschillen in onderzoeksprocedures (= bias) en niet refereert aan een werkelijk verschil tussen de landen. Deze externe validiteit van de resultaten is, zeker in toegepast marktonderzoek, zeer belangrijk.

Het bepalen van bronnen van bias en het vervolgens vaststellen van het bereikte niveau van equivalentie zijn stappen op weg naar beter internationaal vergelijkend onderzoek. Het in dit artikel geschetste raamwerk voor detecteren van bias in het onderzoeksproces en het vaststellen van equivalentie daarna kan hierbij helpen. Uit dit verhaal blijkt dat het vaststellen van equivalentie geen eenvoudige procedure is voor marktonderzoekers in de praktijk. Echter, het bepalen van equivalentie zal het inzicht in verschillen en overeenkomsten tussen landen vergroten en daarmee de kwaliteit van advies gebaseerd op marktonderzoek vergroten.

LITERATUUR

- Aulakh, P.S. & M. Kotabe (1993), "An assessment of theoretical and methodological development in international marketing: 1980-1990," *Journal of International Marketing*, 1(2), 5-28.
- Baumgartner, H. & J.B.E.M. Steenkamp (1996), "Exploratory consumer buying behavior: Conceptualization and measurement," *International Journal of Research in Marketing*, 13(2), 121-137.
- Baumgartner, H. & J.B.E.M. Steenkamp (2001), "Response styles in marketing research: A cross-national investigation", *Journal of Marketing Research*, 38(2), 143-156.
- Bearden, W.O. & R.G. Netemeyer (1999), *Handbook of Marketing Scales*. Second Edition, Sage Publications, Inc Thousand Oaks, California.
- Bryk, A.S., & S.W. Raudenbush (1992), *Hierarchical Linear Models: applications and data analysis methods*. London: Sage.
- Chen, C., S-Y. Lee, & H.W. Stevenson (1995), "Response style cross-cultural comparisons of rating scales among east Asian and north American students," *Psychological Science*, 6(3), 170-175.
- Chinese Culture Connection (1987), "Chinese values and the search for culture-free dimensions of culture," *Journal of Cross-Cultural Psychology*, 18(2), 143-164.

⁴ Dit verwijderen kan, als er veel vragen verwijderd worden, leiden tot (te) kleine schalen en tevens tot andere items die gebiast zijn. Reductie van bias is een werkwijze die met de nodige voorzichtigheid uitgevoerd moet worden.

- Craig, C.S. & S.P. Douglas (2000), *International Marketing Research*. Second Edition, New York: Wiley.
- Dillon, W.R., T.J. Madden, & N.H. Firtle (1994), *Marketing research in a marketing environment*. Third Edition. Irwin.
- Greenleaf, E.A. (1992a), "Improving rating scale measures by detecting and correcting bias components in some response styles," *Journal of Marketing Research*, 29(2), 176-188.
- Hair, J.F., Jr., R.E. Anderson, R.L. Tatham, & W.C. Black (1998). *Multivariate Data Analysis*. Fifth Edition. Upper Saddle River, New Jersey: Prentice-Hall.
- Hays, W.L. (1973), *Statistics for the social sciences*. Second Edition. Holt, London UK: Rinehart and Winston.
- Hofstede, G. (1980), *Culture's Consequences: International differences in Work-Related Values*. Beverly Hills, CA: Sage.
- Holland, P.W. & D.T. Thayer (1988), "Differential item performance and the Mantel-Haenszel procedure," In: *Test Validity*, H. Wainer & H.I. Braun (Eds.) 129-145, Hillsdale NJ: Erlbaum.
- Hui, C.H. & H.C. Triandis (1985a), "Measurement in cross-cultural psychology. A review and comparison of strategies," *Journal of Cross-Cultural Psychology*, 16(2), 131-152.
- Katigbak, M.S., A.T. Church en T.X. Akamine (1996), "Cross-Cultural Generalizability of Personality Dimensions: Relating Indigenous and Imported Dimensions in Two Cultures," *Journal of Personality and Social Psychology*, 70(1), 99-114.
- Kumar, V. (2000) *International marketing research*. Upper Saddle River NJ: Prentice-Hall Inc.
- Landers, S. (1986), "Judge reiterates IQ test ban", *APA Monitor*, 17, 18.
- Lee, C. & R.T. Green (1991), "Cross-cultural examination of the Fishbein behavioral intentions model," *Journal of International Business Studies*, 2, 289-305.
- Lord, F.M. (1980), *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Netemeyer, R.G., S. Durvasula, & D.R. Lichtenstein (1991), "A cross-national assessment of the reliability and validity of the CETSCALE," *Journal of Marketing Research*, vol. XXVIII, 320-327.
- Ryan, A.M, M. Horvath, R.E. Ployhart, N. Schmitt en L.A. Slade (2000), "Hypothesizing Differential Item Functioning in Global Employee Opinion Surveys," *Personnel Psychology*, 53, 531-562.
- Schwartz, S.H. (1992), "Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries," *Advances in Experimental Social Psychology*, 25, 1-65.
- Sekaran, U. (1983), "Methodological and theoretical issues and advancements in cross-national research," *Journal of International Business Studies*, Fall, 61-73.
- Serpell, R. (1979), "How specific are perceptual skills?," *British Journal of Psychology*, 70, 365-380.
- Shimp, T.A. & S. Sharma (1987), "Consumer Ethnocentrism: Construction and Validation of the CETSCALE," *Journal of Marketing Research*, 24(August), 280-289.
- Singh, J. (1995), "Measurement issues in cross-national research," *Journal of International Business Studies*, Third Quarter, 597-619.
- Steenkamp, J.B.E.M. & H. Baumgartner (1998), "Assessing Measurement Invariance in Cross-National Consumer Research," *Journal of Consumer Research*, 25(June), 78-90.
- Ter Hofstede, F., J.B.E.M Steenkamp and M. Wedel (1999), "International Market Segmentation Based on Consumer-Product Relations," *Journal of Marketing Research*, 36(1), 1-17.
- Van der Kloot, W.A. (1997), *Meerdimensionale schaaltechnieken voor gelijkenis- en keuzedata. Ruimtelijke modellen voor psychologie, marktonderzoek en andere wetenschappen*. Utrecht: Lemma.
- Van der Linden, W.J. & R.K. Hambleton (1997), *Handbook of modern item response theory*. Springer-Verlag: New York.
- Van de Vijver, F.J.R. & R.K. Hambleton (1996), "Translating tests: Some practical guidelines," *European Psychologist*, 1(2), 89-99.

- Van de Vijver, F.J.R. & K. Leung (1997a), *Methods and data analysis for cross-cultural research*. Sage.
- Van de Vijver, F.J.R. & K. Leung (1997b), "Methods and data analysis for cross-cultural research," In: *Handbook of Cross-Cultural Psychology, 2nd Ed., Volume I. Theory and Method*, Berry, J.W., Y.H. Poortinga and J. Pandey (Eds.), 257-300. Boston: Allyn and Bacon.
- Van de Vijver, F.J.R. & Y.H. Poortinga (1997), "Towards an integrated analysis of bias in cross-cultural assessment," *European Journal of Psychological Assessment*, 13(1), 21-29.
- Van Herk, H., Th.M.M. Verhallen, & J. Barzilay (1994), "Methodological Issues in international segmentation," In: *Marketing: Its dynamics and challenges*. Bloemer, J., J. Lemmink and J. Kasper (Eds.), 1311-1314. Maastricht NL: EMAC.
- Van Herk, H. (2000), *Equivalence in a Cross-national Context: Methodological and Empirical Issues in Marketing Research*, proefschrift, Katholieke Universiteit Brabant, Tilburg.

7. De invloed van Reclame-Inhoud op het Zapedrag tijdens TV reclames

De ontwikkeling van een “Zap” pretest voor TV reclames

J.L.C.M. WOLTMAN ELPERS*

SAMENVATTING

Het wegzappen van TV reclames door consumenten wordt voor adverteerders een steeds groter probleem. Mede daardoor is de effectiviteit van TV reclame over het algemeen laag. Om de zapkans tijdens de duur van één commercial te minimaliseren, is inzicht nodig in de relatie tussen de inhoud van de commercial en het zapedrag. In dit artikel worden veranderingen in twee belangrijke inhoudscomponenten (“bedoelde prettige gevoelens” en “informatie”) van een specifieke TV reclame per seconde gemeten. Vervolgens worden deze metingen gebruikt om de zapkans van deze TV reclame op elke seconde te voorspellen. Op basis van deze uitkomsten kan de adverteerder naar wens per seconde de reclame-inhoud aanpassen om de zapkans tijdens de commercial zo klein mogelijk te houden.

1. INLEIDING

Mede door de opkomst van commerciële TV zenders, maar ook door toename van de tijdsduur per zender waarop TV reclame mag worden uitgezonden, wordt de consument geconfronteerd met steeds meer TV reclame. In 1990 werd er bijvoorbeeld per dag gemiddeld een klein uur TV reclame op Nederlandse zenders uitgezonden en in 1999 kon de consument ongeveer 14,5 uur per dag naar Nederlandse TV reclame kijken. De enorme vergroting van het aanbod heeft echter niet geleid tot een proportionele stijging in kijktijd van TV reclame. Keek een Nederlandse consument in 1990 nog gemiddeld 2.8 minuten per dag naar Nederlandse TV reclame; nu kijkt hij of zij ongeveer 13 minuten per dag naar Nederlandse TV reclames. Uit onderzoek blijkt dat gemiddeld 13 % van de kijkers tijdens een reclameblok naar een ander kanaal schakelt oftewel zapt (Van Meurs 1999). Het is duidelijk dat de effectiviteit van een specifieke TV reclame verlaagd wordt als veel consumenten zappen tijdens deze commercial of tijdens eerdere commercials in hetzelfde reclameblok.

* De auteur wil graag Michel Wedel en Rik Pieters bedanken voor hun grote bijdragen aan dit onderzoek. Dominique Claessen van Verity International voor het verkrijgen van de Zapping-data. Hans Baumgartner voor de BSP monitor en Frank Harleman voor zijn hulp tijdens de dataverzameling.

Om inzicht te verwerven in (de oorzaken van) het zappedrag tijdens Nederlandse commercials heeft SPOT in 1995 een onderzoek geïnitieerd (Van Meurs 1999). Kort samengevat blijkt dat zappedrag het minst voorkomt in programma-onderbrekende reclameblokken op zenders waar de kijkdichtheid groot is. Reclameblokken na jongeren-programma's hebben de kleinste kans weggezapt te worden en indien consumenten gezamenlijk TV kijken (sociaal kijkgedrag), wordt er ook minder snel gezapt. Als belangrijkste oorzaken voor het zappen tijdens TV reclame wordt aangegeven dat zowel de duur van het programma vòòr als de duur van het reclameblok zelf bepalend zijn: hoe langer, des te groter de zapkans in het reclameblok. Als het reclameblok laat op de avond wordt uitgezonden, treedt zappedrag ook eerder op. Indien het verschil in leeftijd tussen de programma's voor en na het reclameblok groot is en/of programma's op andere zenders populair zijn, zullen veel kijkers tevens de voorkeur geven aan een andere zender. Ander onderzoek wijst uit dat zappedrag tijdens TV reclames mede veroorzaakt wordt door persoonlijke factoren zoals de betrokkenheid van de consument bij het geadverteerde product en hoe vaak de consument blootgesteld is aan een reclame-spot (Siddarth en Chattopadhyay 1998). Reclamefactoren zoals de uniekheid van de boodschap, de duur van de commercial en het tijdstip van uitzending zijn in dit onderzoek tevens bepalende factoren voor het zappedrag tijdens zo'n commercial. Zowel het onderzoek van Siddarth en Chattopadyhay (1998) en van Van Meurs (1999) geven ons een goed inzicht in het zappedrag tijdens commercials indien deze reeds op de televisie worden uitgezonden. Als adverteerder zou men echter ook inzicht in het zappedrag willen verkrijgen alvorens de desbetreffende commercial wordt uitgezonden ("copy-testing"). Verschillende conceptversies van een zelfde TV reclame zouden op hun "zapkans" getest moeten worden. Olney et al. (1991) laten in een experiment zien dat reclame-inhoud naast een indirecte invloed op de kijktijd van TV reclames ook een directe invloed op de kijktijd heeft. Indien de commercial feiten communiceert oftewel informatie over het product/dienst of gebruik ervan geeft, zapt de consument de TV reclame eerder weg dan wanneer de commercial deze informatie niet geeft. Een TV reclame die probeert op een positief gevoel in te spelen wordt langer bekeken. Hetzelfde geldt voor een commercial die uniek is in zijn soort, maar niet te afwijkend is van andere commercials. Naast de directe invloed van deze drie dimensies van reclame-inhoud: "informatie", "bedoelde prettige gevoelens" en "uniekheid" op het zappedrag bepalen deze factoren ook het gevoel van de consument zelf. Een blijde unieke boodschap geeft de consument een prettig gevoel en bevordert een positieve attitude ten aanzien van de commercial, waardoor het zappedrag van deze consument tijdens de commercial verminderd wordt. Het in dit artikel gepresenteerde onderzoek bouwt voort op de studie van Olney et al. (1991) en onderzoekt de relatie tussen de reclame-inhoud en het zappedrag in verder detail.

Reclame-inhoud wordt door de adverteerder beïnvloed en heeft een direct en een indirect effect via emoties en attitudes op het zappedrag. Om het zappedrag tijdens een (concept) TV reclame te beoordelen zal eerst naar de invloed van de verandering in inhoud op het zappedrag gekeken moeten worden. Aangezien tot nu toe weinig bekend is over deze relatie, gaan we in dit artikel hier verder op in. Het hoofddoel is om te bekijken hoe veranderingen in de door de adverteerder opgewekte (on)prettige gevoelens en verstrekte informatie de zapkans tijdens deze commercial veranderen. De uiteindelijke praktische bijdrage is het ontwikkelen van een instrument waarmee een adverteerder inzicht krijgt in het effect van de reclame-inhoud op de zapkans tijdens de commercial. Hiermee kunnen aanbevelingen gedaan worden voor inhoudsveranderingen

gen waarmee het zapedrag tijdens deze (concept-) commercial geminimaliseerd kan worden.

In de volgende paragraaf geven we aan waarom we veranderingen in reclame-inhoud willen meten. Reclame-inhoud delen we in twee belangrijke componenten in: (door de adverteerder) “bedoelde prettige gevoelens” en “informatie”. In deze paragraaf wordt een drietal hypothesen ontwikkeld met betrekking tot de invloed van “bedoelde prettige gevoelens” en “informatie” op de zapkans tijdens TV reclame. Paragraaf 3 en 4 beschrijven de dataverzameling- en analysemethoden en vervolgens worden in paragraaf 5 de resultaten gepresenteerd. Het artikel eindigt met conclusies, discussie en aanbevelingen voor toekomstig onderzoek.

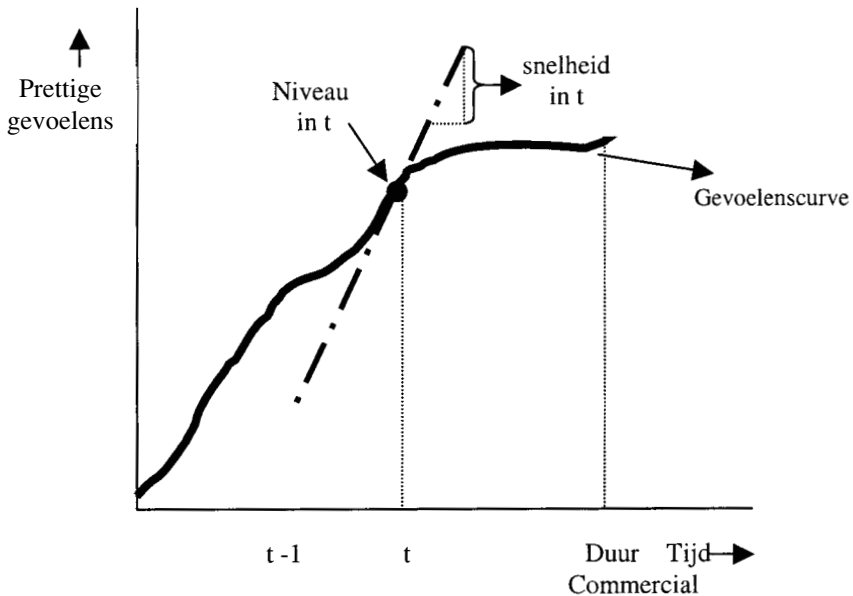
2. DE INVLOED VAN “MOMENT-TO-MOMENT” RECLAME-INHOUD OP ZAPGEDRAG

Baumgartner et al. (1997) laten zien dat het overall gevoel van een consument ten aanzien van een bepaalde commercial niet gelijk is aan het gemiddelde gevoel dat de consument tijdens deze commercial heeft. Het maximale positieve (of negatieve) gevoel tijdens de commercial, het gevoel aan het eind van de commercial en de snelheid waarmee het gevoel verandert tijdens de commercial zijn de bepalende factoren voor de “overall liking” van deze commercial. Baumgartner en zijn collega’s laten zien dat het gevoel tijdens een commercial zeer grote fluctuaties vertoont. Verwacht wordt dat gevoelens en dus mede het zapedrag tijdens de commercial veranderen vanwege veranderingen in de reclame-inhoud. Om dit te onderzoeken moeten we reclame-inhoud dus continu tijdens de duur van de commercial meten. In aansluiting op de bestaande literatuur meten wij de reclame-inhoud op twee dimensies (Aaker et al. 1986; Biel 1998; Puto en Wells 1984): 1) “bedoelde prettige gevoelens” en 2) “informatie”. Met (door de adverteerder) “bedoelde prettige gevoelens” wordt gekeken naar de prettige gevoelens die de reclamemaker met de commercial wil communiceren. Dit wil nog niet zeggen dat de consument deze gevoelens inderdaad ook ervaart. “Informatie” betreft feiten die de commercial communiceert over het product of dienst met betrekking tot de prijs, kwaliteit, prestaties, eigenschappen en onderdelen, smaak, verpakking en vorm, aanbiedingen, garantie, vernieuwingen enzovoort (Resnik en Stern 1977). In de volgende paragraaf wordt uitgelegd hoe we deze twee inhoudscomponenten “moment-to-moment”¹ meten.

We stellen ons in deze paragraaf voor dat het mogelijk is om voor een bepaalde TV reclame “bedoelde prettige gevoelens” en “informatie” continu (moment-to-moment) te meten. Figuur 1 laat de prettige gevoelens² tijdens de duur van deze ene commercial zien. We definiëren de afgebeelde curve als de prettige gevoelenscurve. De waarde van de curve op tijdstip t noemen we het prettige gevoelensniveau op t en de prettige gevoelens-snelheid op tijdstip t is gedefinieerd als de helling van de gevoelenscurve in t . Dezelfde definities worden gehanteerd voor de informatiecurve.

¹ Met “moment-to-moment” reclame-inhoud wordt de reclame-inhoud op elke tijdsmoment in de commercial bedoeld.

² In de rest van het artikel wordt met “prettige gevoelens” expliciet “bedoelde prettige gevoelens” als onderdeel van de reclame-inhoud bedoeld, tenzij anders aangegeven.



Figuur 1. Uitleg van het niveau en de snelheid van de prettige gevoelenscurve.

In het algemeen houden consumenten van prettige gebeurtenissen en worden commercials die een positief gevoel communiceren geprefereerd. Ook Olney et al. (1991) concluderen dat consumenten langer kijken naar “prettige” commercials. We verwachten echter niet alleen dat het niveau van bedoelde prettige gevoelens de zapkans beïnvloedt, maar ook dat snelheid op ieder moment in de gevoelenscurve van de commercial effect heeft op de zapkans: Consumenten kunnen aan een bepaald gevoelensniveau gewend raken en houden van verandering in dit gevoelensniveau in positieve richting (Frederick en Loewenstein 1999). Baumgartner et al. (1997) concluderen dat een positieve snelheid in prettige gevoelens tijdens de TV reclame de “overall liking” van de commercial laat stijgen. We verwachten mede daarom dat een positieve snelheid in de bedoelde prettige gevoelenscurve de zapkans laat dalen. Dus, gebaseerd op bovenstaande bevindingen ontwikkelen we de volgende hypothese:

H₁: Een hoog niveau en een positieve snelheid in de bedoelde prettige gevoelenscurve laten de zapkans tijdens de TV reclame dalen.

Onderzoek naar printadvertenties concludeert dat consumenten langer kijken naar advertenties die veel informatie bevatten (Pieters, Warlup en Wedel 2001). We verwachten echter een ander effect bij TV reclames die veel informatie bevatten. Terwijl consumenten de informatie in printadvertenties op hun eigen tempo kunnen bekijken en verwerken, wordt de hoeveelheid informatie en de informatie-verwerkingssnelheid in TV commercials door de reclame zelf bepaald. Tegelijkertijd gaan consumenten tijdens het kijken naar TV reclames niet expliciet op zoek naar informatie en worden ze liever door de TV reclame geamuseerd (passief kijkgedrag; Chaiken en Eagly 1983). Pasadeos (1990) laat tevens zien dat een hoge informatieniveau in TV- en radiocommercials de

irritatie van consumenten doet toenemen. Daarom verwachten we dat een hoog informatieniveau in TV reclames de kijktijd negatief beïnvloedt (Olney et al. 1991). Hetzelfde wordt verwacht voor de snelheid in de informatieverschaffing: Indien een hoog informatieniveau al niet wordt geprefereerd, willen consumenten ook niet dat de hoeveelheid informatie toeneemt in de TV commercial (positieve snelheid). We voorspellen daarom:

H₂: Een hoog niveau en een positieve snelheid in de informatiecurve laten de zapkans tijdens de TV reclame stijgen.

Indien adverteerders de zapkans tijdens de commercial zo klein mogelijk willen maken, zouden bovenstaande hypothesen kunnen impliceren dat hoge niveaus en snelheden in de informatiecurve gecompenseerd kunnen worden door voldoende hoge niveaus en snelheden in de prettige gevoelenscurve. Echter, één onderzoeksstroom in de psychologie beargumenteert dat prettige gebeurtenissen en gemoedstoestanden de motivatie en capaciteit om informatie verwerken reduceren (Schwartz 2001). Zulke gebeurtenissen en gemoedstoestanden kunnen een signaal aan de persoon afgeven dat er geen gevaar dreigt en er dus minder aandacht besteed hoeft te worden aan externe informatie. Ook in studies van MacKie en Worth (1991) blijken mensen in een prettige gemoedstoestand meer tijd nodig te hebben om bepaalde teksten te lezen. De auteurs beargumenteren dat plezierige gebeurtenissen en gemoedstoestanden de informatieverwerkingscapaciteit reduceren en de persoon dus op hetzelfde moment minder aandacht aan externe informatie kan besteden. In TV reclame wordt de informatieverwerkingsnelheid door de commercial zelf bepaald: Indien de informatieverwerkingsmotivatie of -capaciteit van de consument door een hoog prettig gevoelensniveau beperkt wordt, kan een hoog informatieniveau in de TV reclame de consument teveel belasten. De consument zal deze belasting willen ontwijken (Jacoby 1984) en zappedrag is het gevolg.

Eén andere onderzoeksstroming binnen psychologie onderscheidt twee verschillende informatieverwerkingsstrategieën, namelijk: 1) gebaseerd op interne informatie dat reeds binnen de hersenen aanwezig is en 2) gebaseerd op externe informatie die de omgeving aan de persoon verschaft (Bless 2000; Fiedler 2000). Indien de consument zich goed voelt (door bijvoorbeeld een prettige gebeurtenis), besteedt de consument meer aandacht aan informatie die reeds in de hersenen aanwezig is. Bestaande kennisstructuren worden door het positieve gevoel geactiveerd en de aandacht voor externe informatie van bijvoorbeeld een TV reclame zal sterk verminderen of verdwijnen. Consumenten die zich op een bepaald moment slecht voelen daarentegen hebben juist wel meer aandacht voor externe informatie. Indien een consument tijdens het kijken naar TV reclame veel bedoelde prettige gevoelens en informatie krijgt te verwerken, stimuleren de bedoelde prettige gevoelens aandacht te besteden aan reeds opgeslagen informatie in de hersenen, maar de informatie in de reclame motiveert juist een andere informatieverwerkingsstrategie, namelijk om meer aandacht te besteden aan de informatie in de commercial. Dit kan een conflictsituatie opleveren. Omdat consumenten geen controle hebben over de informatiehoeveelheid en -snelheid in de TV reclame, is zappen een manier om deze conflictsituatie te vermijden. Gebaseerd op bovenstaande redenties, voorspellen we dat:

H₃: Een hoog niveau en positieve snelheid in de bedoelde prettige gevoelenscurve in combinatie met een hoog niveau en positieve snelheid in de informatiecurve laten de zapkans tijdens de TV reclame stijgen.

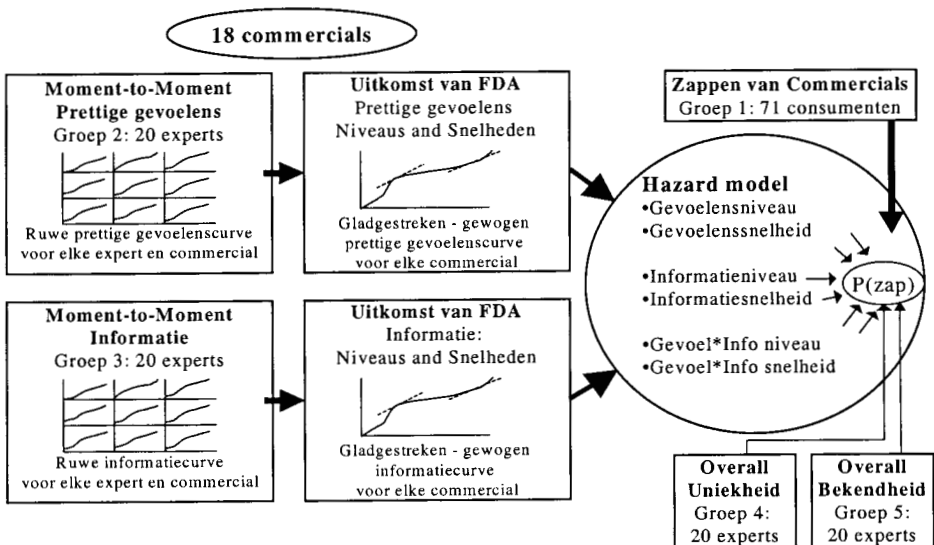
3. DATAVERZAMELING

Voordat we uitleggen hoe we de veronderstelde relaties onderzoeken geven we een stroomdiagram in figuur 2 waarin we aangeven:

hoe de data worden verzameld en geanalyseerd. Om voor specifieke commercials de beste mogelijke objectieve continue beoordeling van de twee inhoudscomponenten te krijgen, maken we gebruik van inhoudsanalyse.

Twee onafhankelijke groepen van elk 20 getrainde experts beoordelen 18 TV reclames op “moment-to-moment” bedoelde prettige gevoelens en informatie. Op basis van de beoordelingen van de experts wordt per inhoudscomponent de “beste” (intersubjectieve) maatstaf van deze component voor elke commercial verkregen door middel van functionele data analyse (FDA). De methode wordt in paragraaf 4 besproken. Om in het model te corrigeren voor de effecten van reclame uniekheid en reclamebekendheid geven twee andere groepen van experts per commercial overall beoordelingen van deze twee maatstaven. Van een steekproef van 71 consumenten worden zaptijden van dezelfde commercials verkregen.

Men zou kunnen beargumenteren dat het de voorkeur heeft de metingen van de reclame inhoud en het zapedrag van één steekproef consumenten te verkrijgen. Er zijn echter verschillende redenen waarom we voor een design van onafhankelijke steekproeven van experts en consumenten hebben gekozen. Het is ten eerste bijna onmogelijk aan een consument te vragen tegelijkertijd de reclame continu zowel op “bedoelde prettige gevoelens” als op “informatie” te beoordelen en het gewenste moment van zappen aan te geven. Indien we deze drie taken zouden splitsen en de consument de reclame dus voor elke taak opnieuw moet zien, kunnen ongewenste herhalings-effecten optreden. Tenslotte zijn we geïnteresseerd in de door de adverteerder *bedoelde* reclame-inhoud en experts kunnen dit beter beoordelen dan consumenten.



Figuur 2. Stroomdiagram van dataverzameling en -analyse.

3.1. Stimuli

In deze studie worden 18 TV reclames gebruikt, die op de Nederlandse televisie zijn uitgezonden. De commercials vertegenwoordigen de volgende productcategorieën: frisdranken, alcoholische dranken, bankdiensten, wasmiddelen, scheer producten, maandverband, vliegreizen, snacks, zeeproducten, telecommunicatie, supermarkten, auto's en koffie. De duur van de commercials varieert van 20 tot 45 seconden.

3.2. Zapping data

Zapedrag wordt in deze studie geoperationaliseerd als het beëindigen van het zien van een commercial door met de afstandsbediening naar een andere zender te schakelen. 71 consumenten (32 mannen en 39 vrouwen in de leeftijd van 18 tot 60 jaar) namen deel aan een groot adverteerdersonderzoek. De 18 commercials werden in 5 blokken aan de respondenten met een afstandsbediening in de hand getoond. De instructie was: "U gaat straks verschillende TV programma's en commercials zien. Als u een programma of TV commercial wilt zappen, drukt u op de daarvoor bestemde knop". Indien de consument de commercial zapte, werd een andere commercial of TV programma getoond om een zo'n natuurlijke mogelijke situatie na te bootsen. Commercials werden in verschillende volgordes aan respondenten getoond. Tussen de reclameblokken hadden respondenten een pauze of voerden taken uit die niet gerelateerd waren aan het TV reclame onderzoek.

3.3. "Moment-to-moment" prettige gevoelens

20 getrainde experts (vierde jaars studenten met specialisatie in reclame- en communicatieonderzoek, 10 mannen en 10 vrouwen) evalueerden de 18 commercials verdeeld over 5 reclameblokken op "bedoelde prettige gevoelens". We gebruiken hiervoor een instrument dat per seconde het aangegeven niveau in gevoelens registreert (BSP monitor). De expert krijgt een computerscherm voor zich waarop een schaal is afgebeeld. Bij bedoelde gevoelens staat aan de linkerkant "helemaal niet prettig" en aan de rechterkant "heel erg prettig" en in het midden van de schaal wordt een neutraal punt aangegeven. De experts krijgen de opdracht de criteria zo te interpreteren als in paragraaf 2 is aangegeven en de bedoelde en niet de gevoelde inhoud te evalueren. Zodra de commercial op een televisiescherm begint, start op het computerscherm een lijn over het neutrale punt van de schaal naar beneden te lopen. Zodra de expert denkt dat de commercial prettige gevoelens wil communiceren, beweegt hij of zij de muis naar rechts en zodra hij of zij denkt dat de commercial onprettig wil zijn, wordt de muis naar links bewogen. De expert kan het verloop van de lijn op het scherm volgen. De lijn verdwijnt van het beeld, zodra de commercial eindigt. De computer registreert automatisch per seconde het niveau van de bedoelde prettige gevoelens. Baumgartner et al. (1997) gebruikt dit instrument (BSP monitor) voor registratie van affectieve reacties ten aanzien van TV reclames en VandenAbeelee en MacLachlan (1994) hebben de betrouwbaarheid en validiteit van dit instrument onderzocht³. Om mogelijke carry-over effecten te voorkomen werd na elke commercial een animatiefilmpje getoond en kreeg elke expert de commercials in een andere volgorde te zien (VandenAbeelee en MacLachlan 1994). Tussen de 5 reclameblokken hielden de experts een pauze.

³ Zie voor verdere details in Woltman Elpers et al. (2001), verkrijgbaar bij de auteur.

3.4. “Moment-to-moment” informatie

20 andere getrainde experts (vierde jaars studenten met specialisatie in reclame- en communicatieonderzoek, 10 mannen en 10 vrouwen) beoordeelden de 18 TV reclames op hun informatiegehalte. Een zelfde procedure als hierboven werd gehanteerd, waarbij een schaal van “helemaal niet informatief” tot “heel erg informatief” werd gebruikt. Benadrukt werd dat experts de commercial op aanwezigheid van één of meerdere informatiekenmerken zoals in paragraaf 2 beschreven moesten beoordelen.

3.5. Commercial uniekheid en bekendheid

Om te kunnen corrigeren voor effecten van twee andere belangrijke variabelen in het model, namelijk commercial uniekheid en reclamebekendheid (Pieters et al. 2001; Olney et al. 1991; Siddarth en Chattopadhyay 1998) beoordeelden 2 verschillende groepen van experts (elke vierde jaars studenten met specialisatie in reclame- en communicatieonderzoek, 10 mannen en 10 vrouwen) de 18 commercials op uniekheid en bekendheid. Voor uniekheid worden 5 items op 7-puntsschalen beoordeeld (1 = helemaal niet..., 7 = heel erg ...), namelijk: “afwijkend van andere commercials”, “nieuw”, “gewoon”, “verrassend” en “origineel”. Data met betrekking tot de bekendheid van de Nederlandse bevolking met de commercials en geadverteerde merken waren op het tijdstip van dit onderzoek niet beschikbaar. Daarom is hiervoor gebruikt gemaakt van beoordeling van experts. 20 experts beoordeelden eerst een lijst met merknamen (inclusief de geadverteerde merken, niet geadverteerde merken en niet bestaande (controle-) merken) op bekendheid van de Nederlandse bevolking in het algemeen met deze merken op 7-puntsschalen (1 = helemaal niet bekend, 7 = heel erg bekend). Hetzelfde werd gedaan voor een lijst met productcategorieën. Hierna kregen de experts de 18 commercials te zien en onmiddellijk na elke commercial werd gevraagd hoe bekend de Nederlandse bevolking met de commercial zou zijn. Alle Cronbach’s alpha’s zijn groter dan 0.70. Als maat voor uniekheid en bekendheid wordt per commercial het gemiddelde over de bij de dimensie behorende items en over alle experts genomen.

4. DATA ANALYSE

4.1. FDA voor het verkrijgen van een prettige gevoelenscurve en informatiecurve per commercial

In het onderstaande behandelen wij eerst de FDA analyse om één prettige gevoelenscurve per commercial te verkrijgen. Dezelfde procedure is toegepast voor het verkrijgen van één informatiecurve per TV reclame.

Om één prettig gevoelenscurve per commercial te verkrijgen, kan men per seconde over de 20 individuele curven middelen (Baumgartner et al. 1997). Van de verkregen prettige gevoelenscurve zou men als maatstaf voor de snelheid op elke seconde t het verschil in prettige gevoelens tussen twee opeenvolgende seconden kunnen nemen. Te verwachten is echter dat de ene expert beter is dan de andere expert in het vinden van de “beste” prettige gevoelenscurve voor deze commercial. Indien we het gemiddelde nemen over alle experts nemen we echter aan dat elke expert even goed is in het vinden van deze curve. Tevens blijkt het verschil in prettige gevoelens tussen twee opeenvol-

gende seconden een slordige maatstaf te zijn voor de eerste orde afgeleide (dit is namelijk de snelheid; Ramsey en Silverman 1997). Daarom gebruiken we een functionele data analyse methode (FDA, zie vb. Ramsey en Silverman 1997) om een betere representatie van de prettige gevoelenscurve per commercial te vinden. In ons geval komt deze analyse op het volgende neer³: Wij gaan er van uit dat elke expert de prettige gevoelensfunctie van iedere commercial over de tijd zo goed mogelijk probeert terug te vinden. De ene expert doet dat beter dan de andere. De “beste” prettige gevoelensfunctie kennen we niet, anders zouden we de experts ook niet nodig hebben. Eerst proberen we voor elke expert per commercial een functionele vorm te verkrijgen. We gebruiken hiervoor een non-parametrische techniek, namelijk lokale kwadratische regressie. Na enig analysewerk hebben we dan voor de gevoelensbeoordelingen van iedere expert per commercial een functionele vorm. Het is nu per expert mogelijk om voor elke seconde op een verantwoorde manier de eerste orde afgeleide te berekenen. Uiteindelijk krijgen we voor elke commercial een matrix van de 20 prettige gevoelensfuncties met in de rijen de seconden en in de kolommen de experts en een corresponderende matrix van de eerste orde afgeleiden per seconde voor de 20 experts. Indien we vervolgens de matrices van de prettige gevoelensfuncties rijgewijs poolen en een principale componenten analyse (PCA) op de covariantie matrix van de gepoolde matrix uitvoeren en de eerste component bewaren, verkrijgen we gewichten voor elke expert, zodanig dat de variantie over experts gemaximaliseerd wordt. Indien elk gewicht wordt vermenigvuldigd met de bijbehorende individuele curve krijgen we per commercial een gewogen gemiddelde prettige gevoelenscurve, hetgeen ook het doel van de analyse is. De eerste orde afgeleiden behorende bij deze “beste” prettige gevoelenscurve worden verkregen door de matrix van eerste orde afgeleiden met de verkregen gewichtenvector uit de PCA analyse te vermenigvuldigen. Door middel van FDA verkrijgen we uiteindelijk per commercial een prettige gevoelens- en informatiecurve met bijbehorende snelheden in elk tijdstip t .

4.2. Hazard model voor het modeleren van de zapkans

We schatten een random effects discrete tijd hazard model voor de zapkans. De afhankelijke variabele (hazard) in dit model wordt geïnterpreteerd als de zapkans op tijdstip t (hier seconde t). Stel dat $i = 1, \dots, I$ de commercials voorstelt, $p = 1, \dots, P$ de consumenten, $q = 1, \dots, Q_{ip}$ het aantal tijdsintervallen waarin de kijktijd van consument p voor commercial i is opgedeeld en t_{ip} de tijdsduur gedurende consument p blootgesteld is aan commercial i . X_{iq} geeft de matrix met tijdsafhankelijke onafhankelijke variabelen aan (prettige gevoelens- en informatieniveaus en snelheden) en Z_i stelt de matrix voor met covariaten die constant over de tijd blijven (uniekheid en bekendheid).

We schatten een “piece-wise” exponentiële formule voor de zapkansfunctie (hazard) en we verdelen de kijktijd van elke consument p voor commercial i op in Q_{ip} discrete tijdsintervallen van één seconde. Dit zorgt ervoor dat de zapkans constant is binnen elke seconde en de tijd in deze seconde exponentieel verdeeld is. Zapmomenten worden toegeschreven aan het middenpunt van elke seconde. De kans dat commercial i met kenmerken X_{it} en Z_i gezapt wordt door consument p in de q -de seconde na het begin is dan:

$$(1) \quad h_{ipq} = h_{ip}(t_q) * h_1(x_{iq}) * h_2(z_i)$$

met een baseline zapkans (zapkans die alleen afhankelijk is van de tijd)

$$(2) \quad h_{ip}(t_q) = \exp(\alpha_{0ip} + \alpha_1 t_q + \alpha_2 t_q^2 + \alpha_3 \ln t_q)$$

met $\alpha_{0ip} \sim N(\bar{\alpha}_0, \sigma_{\alpha_0}^2)$ en

$$(3) \quad h_1(X_{iq}) = \exp(\beta_1 \bar{a}_{iq} + \beta_2 \bar{c}_{iq} + \beta_3 \bar{a}_{iq} \bar{c}_{iq} + \beta_4 \bar{a}_{iq}^{(1)} + \beta_5 \bar{c}_{iq}^{(1)} + \beta_6 \bar{a}_{iq}^{(1)} \bar{c}_{iq}^{(1)})$$

$$(4) \quad h_2(Z_i) = \exp(\gamma_1 u_i + \gamma_2 f_i)$$

met prettig gevoelensniveau \bar{a}_{iq} en -snelheid $\bar{a}_{iq}^{(1)}$ en met informatieniveau \bar{c}_{iq} en -snelheid $\bar{c}_{iq}^{(1)}$ van commercial i gedurende tijdsinterval q . De interactie tussen het prettige gevoelens- en informatieniveau van commercial i gedurende tijdsinterval q wordt aangegeven door $\bar{a}_{iq} \bar{c}_{iq}$, de interactie tussen de snelheden door $\bar{a}_{iq}^{(1)} \bar{c}_{iq}^{(1)}$. De uniekheid van commercial i wordt aangegeven door u_i en de bekendheid van TV reclame i door f_i .

Om rekening te houden met heterogeniteit in de gemiddelde zapkans voor verschillende commercials en consument is een random intercept $\alpha_{0ip} \sim N(\bar{\alpha}_0, \sigma_{\alpha_0})$ in het model opgenomen. De baseline zapkans in vergelijking (2) heeft een flexibele vorm, zodat de kans op een verkeerd gespecificeerd model geminimaliseerd wordt (Trussel and Richards 1985). De parameters van het hazard model worden geschat door het maximaliseren van de gesimuleerde likelihoodfunctie³. De geschatte parameters van de onafhankelijke variabelen zijn net zo te interpreteren als regressiecoëfficiënten.

5. RESULTATEN

Eerst bespreken we de resultaten van de FDA analyses die zijn verricht voor het verkrijgen van representatieve prettige gevoelens- en informatiecurven met bijbehorende snelheden per TV reclame. Vervolgens kijken we naar de resultaten van het hazard model.

5.1. Resultaten van de FDA analyses

Voor zowel “informatie” als “bedoelde prettige gevoelens” zijn de beoordelingen over de 20 experts consistent. De mediaan Cronbach’s alpha voor “bedoelde prettige gevoelens” is gelijk aan 0.96 en voor “informatie” is deze gelijk aan 0.83. Daaruit mogen we concluderen dat de meningen van experts over de bedoelde prettige gevoelens en informatie per seconde in voldoende mate overeenkomen. Eerst maken we van de 20 gevoelens (informatie) curven 20 individuele gevoelens (informatie) functies met behulp van lokale kwadratische regressie en berekenen van iedere curve de eerste orde afgeleiden. PCA op de gepoolde 20 gevoelensfuncties levert een eerste component op die 63 % van de variantie van de beoordelingen van de experts verklaart. PCA op de informatiefuncties levert een eerste component op die 56 % van de variantie over de experts verklaart. De gewichten van de experts voor de representatieve gevoelenscurve lopen uiteen van 0.026 tot 0.087 (som van gewichten is gelijk aan 1), voor de informatiecurven liggen deze gewichten tussen de 0.001 en 0.264. Geconcludeerd mag worden dat experts in voldoende mate overeenkomen in hun beoordelingen van de twee inhoudscomponenten. Experts komen echter meer in hun beoordelingen van prettige bedoelde

gevoelens overeen en verschillen meer in hun beoordelingen over het informatiegehalte van de commercials. Blijkbaar is het moeilijker het informatiegehalte intersubjectief te bepalen. Uit onze resultaten blijkt dat bij toepassing van FDA beter rekening wordt gehouden met individuele verschillen tussen experts dan bij het berekenen van een ongewogen gemiddelde curve. Ook kunnen met behulp van onze methode goede metingen van de eerste orde afgeleiden worden verkregen.

5.2. Resultaten van het hazard model

Indien we het hazard model schatten worden de volgende resultaten verkregen:

Tabel 1. Geschatte parameters in het hazard model.

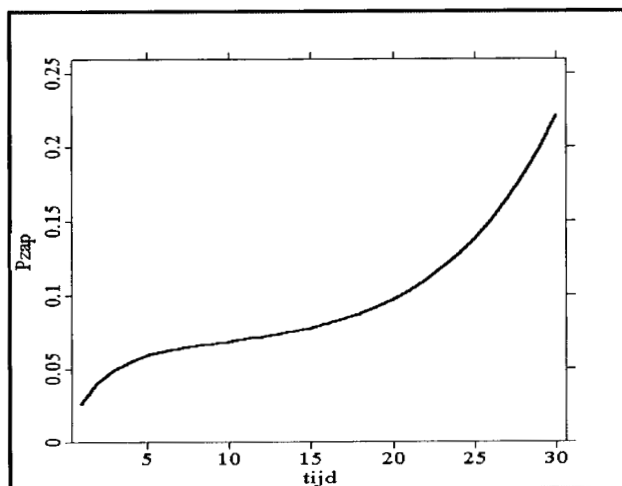
Variabele	Geschatte parameter	P-waarde
Gemiddelde constante	-3.353	<.001
Spreiding van de constante	0.088	.273
Tijd	-1.338	<.001
Tijd in kwadraat	1.355	<.001
Natuurlijke logaritme van tijd	0.754	<.001
Prettig gevoelensniveau	-0.164	<.001
Informatieniveau	0.264	<.001
Prettige Gevoelens*Informatieniveau	0.385	<.001
Prettige gevoelensnelheid	-0.261	<.001
Informatiesnelheid	0.317	<.001
Prettige Gevoelens*Informatiesnelheid	0.126	<.001
Uniekheid	-0.118	0.005
Bekendheid	0.001	0.4041

De gemiddelde constante en de spreiding van de constante zijn opgenomen in het model om rekening te houden met verschillen tussen personen en commercials. Verschillen tussen personen en commercials zijn niet significant. Onafhankelijk van tijds-effecten en andere onafhankelijke variabelen wordt gemiddeld 3.4 % van de commercials in elke seconde weggezapt⁴. Indien we de zapkans gebaseerd op het aantal resterende kijkers tegen de tijd afzetten voor een 30 seconden commercial zoals in figuur 3, zien we dat de zapkans stijgt, zodra de commercial langer duurt. In het begin stijgt de zapkans zeer sterk, na 5 seconden stijgt de zapkans nog enigszins en de stijging neemt weer toe tot het einde van de commercial. Blijkbaar zappen veel mensen de commercial in het begin en van het aantal mensen die de commercial blijven bekijken neemt het aantal zappers per seconde steeds meer toe.

Het negatieve teken van uniekheid geeft aan dat hoe unieker de commercial is, des te kleiner de zapkans. Dit bevestigt de resultaten van Siddarth en Chattopadyhay (1998). Bekendheid met de commercial heeft in onze studie geen effect op de zapkans. Alle hypothesen worden door de resultaten ondersteund. Een hoog prettig gevoelensniveau laat de zapkans dalen (-1.64, P<.001) en een hoog informatieniveau laat de zapkans stijgen (0.264, P<.001). Zodra de commercial toeneemt in het prettige gevoelens-

⁴ Dit getal is als volgt verkregen: $(\exp(-3.353)/(1+\exp(-3.353))) * 100\%$.

niveau (positieve snelheid), neemt de zapkans af ($-0.261, P < .001$). Indien de commercial echter toeneemt in het geven van informatie (positieve snelheid), neemt de zapkans toe ($0.317, P < .001$). Consumenten blijken te houden van snelle prettige gevoelige commercials, maar prefereren geen grote toename in informatie.



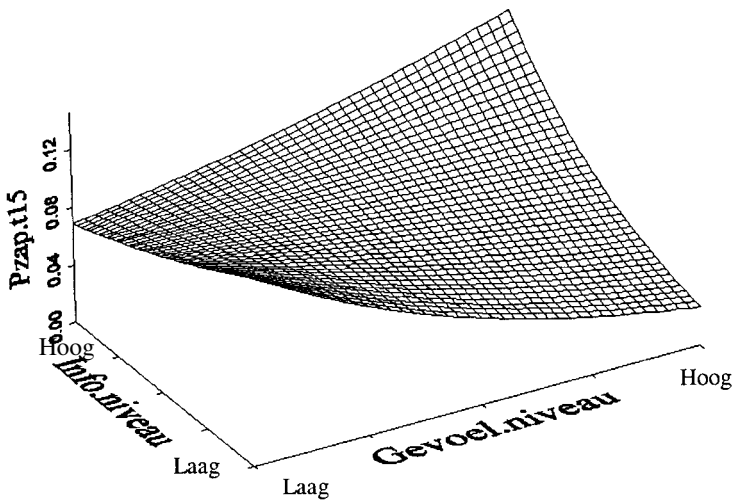
Figuur 3: De baseline zapkans gedurende een commercial van 30 seconden.

Zoals we in hypothese 3 verwachtten, hebben de niveaus en snelheden in prettige gevoelens en informatie een sterk gezamenlijk effect. Het positieve coëfficiënt van “Prettige Gevoelens*Informatie niveau” laten zien dat hoge niveaus in beide inhoudscomponenten op hetzelfde moment de zapkans laten toenemen ($0.385, P < .001$). Ditzelfde geldt voor de interactieterm van de snelheden ($0.126, P < .001$). Om het effect van de interactieterm van de niveaus op de zapkans gedetailleerder te onderzoeken, illustreert figuur 4 een 3-dimensionale grafiek met het prettige gevoelens- en informatieniveaus op de x en y -as en de geschatte zapkans op de z -as voor $t = 15$ seconden.

Zoals deze figuur laat zien, is de zapkans maximaal, als zowel het prettige gevoelens- als informatieniveau maximaal zijn. De zapkans is het laagst als het prettige gevoelensniveau hoog en het informatieniveau laag is. Als het prettige gevoelensniveau laag is, ligt de zapkans tussen deze twee extreme waarden. Deze resultaten ondersteunen onze redenering in paragraaf 2: Indien men zoveel mogelijk kijkers tijdens een TV reclame wil behouden, heeft in een bepaalde seconde een hoog prettig gevoelensniveau in combinatie met een laag informatieniveau de voorkeur boven hoge niveaus in beide inhoudscomponenten. De interactie term voor de snelheden wordt op dezelfde manier geïnterpreteerd. Deze effecten blijven hetzelfde in eerdere en latere seconden in de commercial. Het enige verschil is dat alle zapkansen hoger (lager) liggen in dit latere (eerdere) tijdstip.

5.3. Praktische toepassing

Het hierna volgende voorbeeld illustreert hoe onze methodologie gebruikt kan worden



Figuur 4: Illustratie van de gezamenlijke invloed van prettige gevoelens- en informatieniveaus op de zapkans.

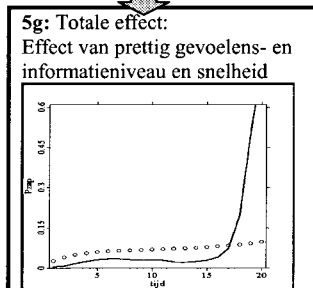
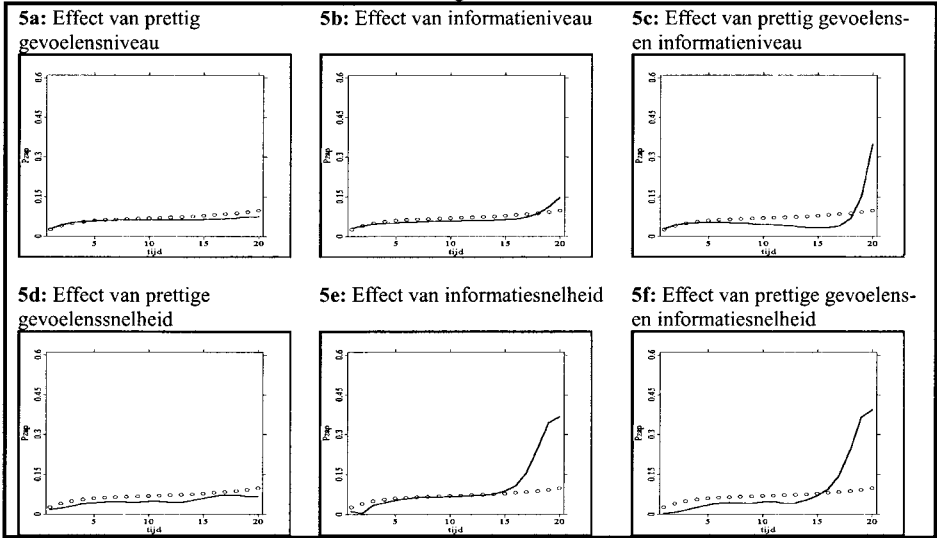
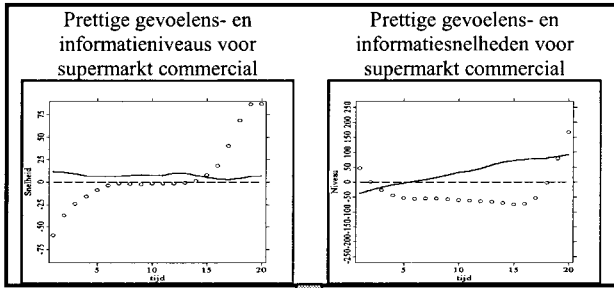
N.B. x-as is het informatieniveau (van achteren naar voren is van hoog naar laag), y-as is het prettige gevoelensniveau (van links naar rechts is van laag naar hoog), z-as is de zapkans (van beneden naar boven is van laag naar hoog).

door adverteerders. We nemen bijvoorbeeld een commercial voor een supermarkt die 20 seconden duurt. De adverteerder laat bijvoorbeeld door een marktonderzoekbureau met behulp van twee groepen van elk 20 experts deze commercial naast andere (concept-) commercials per seconde evalueren op “bedoelde prettige gevoelens” en “informatie”. Met behulp van de door ons voorgestelde FDA analyses is het voor deze adverteerder mogelijk een representatieve curve voor de bedoelde prettige gevoelens en informatie met corresponderende snelheden te verkrijgen, zoals bovenaan in figuur 5 weergegeven is.

De coëfficiënten uit tabel 1 worden gebruikt om de zapkans voor deze TV reclame over de tijd uit te rekenen. Figuren 5a t/m 5g brengen de effecten van de verschillende niveaus en snelheden in beeld. Op de horizontale as staat de tijd en op de verticale as staat de zapkans. De gestippelde lijn laat de baseline zapkans zien voor deze commercial. De zwarte lijn laat de invloed van de desbetreffende onafhankelijke variabele(n) zien.

Het lage prettige gevoelensniveau aan het begin van de commercial doet de zapkans een beetje stijgen, maar het niveau wordt steeds prettiger, zodat de zapkans ook steeds meer daalt (zie figuur 5a). Het hoge informatieniveau aan het begin en eind van de commercial doet de zapkans stijgen in deze tijdsperioden. In het midden van de commercial laat het lage informatieniveau de zapkans dalen (zie figuur 5b). Zoals uit de resultaten blijkt, hebben het prettige gevoelensniveau en het informatieniveau een groot gezamenlijk effect op de zapkans: de zapkans stijgt in het begin een beetje door een laag prettig gevoelens- en informatieniveau. Echter, de zapkans daalt spoedig doordat het gevoelensniveau steeds prettiger wordt en het informatieniveau lager wordt. Aan

N.B.
 Zwarte lijn is
 "gevoelens";
 Stippellijn is
 "informatie"



N.B.
 In figuren 5a t/m 5g geeft de stippellijn de baseline kans aan en geeft de zwarte lijn het effect van de desbetreffende variabele(n) op de baseline zapkans aan.

Figuur 5: De effecten van prettige gevoelens- en informatieniveaus en snelheden op de zapkans in een supermarktcommercial.

het eind van de commercial zijn beide niveaus echter weer hoog met een stijging in de zapkans als gevolg. De toename in de positieve snelheid van de bedoelde prettige gevoelens in de commercial zorgt ervoor dat de zapkans daalt gedurende de duur van deze TV reclame (figuur 5d). De daling in de informatiesnelheid doet de zapkans aan het begin dalen, maar de positieve informatiesnelheid aan het eind van de commercial

laat de zapkans weer stijgen, zoals afgebeeld in figuur 5e. Figuur 5f laat een duidelijk groot gezamenlijk effect zien van beide snelheden op de zapkans, met name aan het einde van de commercial. Het totale effect van alle prettige gevoelens- en informatievariabelen op de zapkans is weergegeven in figuur 5g.

Op de hierboven beschreven wijze kan de adverteerder dus inzicht verwerven in het effect van de twee voornaamste inhoudscomponenten op de zapkans tijdens deze supermarktcommercial. De meeste kijkers zappen weg gedurende de eerste 2 seconden; Van de resterende kijkers, zappen de meeste in de laatste 2 seconden. We weten van de hierboven beschreven analyse dat dit verloop van de zapkans over de tijd veroorzaakt wordt door bepaalde inhoudskenmerken. Door de niveaus en snelheid in de prettige bedoelde gevoelens en informatie op bepaalde momenten aan te passen, kan de zapkans op deze momenten worden veranderd.

6. CONCLUSIES EN DISCUSSIE

6.1. Conclusies

Sinds de afstandsbediening het voor de consument mogelijk maakt TV programma's en TV reclames weg te zappen, wordt onderzoek naar zapedrag tijdens TV reclames steeds belangrijker voor adverteerders en producenten. Andere onderzoek laat zien dat reclame-inhoud de kijktijd naar TV reclames direct kan beïnvloeden (Olney et al. 1991). Tot dusverre is onderzoek naar de invloed van specifieke kenmerken in reclame-inhoud op de zapkans tijdens de duur van de TV reclame achterwege gebleven. Reclame-inhoud verandert gedurende de duur van een commercial en daarom wordt er verwacht dat de zapkans ook mede hierdoor tijdens de uitzending van een TV reclame verandert. In dit artikel wordt getracht een instrument te ontwikkelen waarmee twee belangrijke inhoudscomponenten in verband worden gebracht met de zapkans gedurende een commercial. Deze inhoudscomponenten zijn: "bedoelde prettige gevoelens" en "informatie". Beide inhoudscomponenten kunnen direct door de adverteerder gemanipuleerd worden. Dit artikel doet een eerste aanzet tot het gebruik van een instrument om deze twee inhoudscomponenten voor elke seconde in de commercial te meten. Tevens beschrijven we enkele concepten uit functionele data analyse om op basis van expertbeoordelingen een representatieve maat voor beide inhoudscomponenten te verkrijgen. Er wordt niet alleen gekeken naar de niveaus in "bedoelde prettige gevoelens" en "informatie", maar ook naar de snelheden hiervan. Hoge niveaus en snelheden in "bedoelde prettige gevoelens" bevorderen een positieve attitude met betrekking tot de commercial en dus de motivatie te blijven kijken. Hoge niveaus en snelheden in "informatie" daarentegen laten de zapkans stijgen. Hoge informatieniveaus en -snelheden worden irritant bevonden of belasten de capaciteit van informatieverwerking van consumenten teveel, vooral omdat de commercial de informatiehoeveelheid en -snelheid bepaalt. Onze studie laat echter zien dat een adverteerder bijvoorbeeld verkeerd geadviseerd kan worden het prettige gevoelensniveau van informatiele commercials op te schroeven, aangezien hoge informatieniveaus en -snelheden onverenigbaar zijn met hoge prettige gevoelensniveaus en -snelheden. Dit zou veroorzaakt kunnen worden doordat prettige gevoelens in een commercial de motivatie en/of capaciteit reduceert om informatie te verwerken, zodat een combinatie van prettige gevoelens en veel informatie de consument overbelast en dus zapedrag stimuleert. Een andere verklaring is

dat bedoelde prettige gevoelens stimuleert aandacht te besteden aan reeds opgeslagen informatie in de hersenen en veel informatie in de commercial juist motiveert aandacht te besteden aan deze andere externe informatie. Deze informatie-verwerkingsstrategieën kunnen leiden tot conflictsituaties bij hogere niveaus en snelheden in bedoelde prettige gevoelens en informatie en dus leiden tot zapgedrag.

6.2. Discussie

Deze studie is een eerste poging om de invloed van “moment-to-moment” reclame-inhoud op de zapkans te onderzoeken. Vele studies zullen moeten volgen om de door ons bevonden resultaten te toetsen, voordat er een betrouwbaar en gevalideerd instrument ontwikkeld kan worden dat adverteerders kan helpen de zapkans van commercials door inhoudsanalyse met experts te manipuleren. In onze studie maken wij gebruik van onafhankelijke steekproeven van experts om verschillende inhoudsaspecten van een commercial te beoordelen. We gebruiken deze metingen om consumentengedrag te voorspellen. Als reden voor het gebruik van experts beschrijven we in paragraaf 3 dat het onmogelijk is met de door ons gebruikte instrumenten gelijktijdig metingen van verschillende commercial aspecten te verkrijgen van één respondent. Dit zou echter wel kunnen indien we instrumenten zoals hersenscans of huidreacties zouden kunnen gebruiken om bepaalde emoties bij consumenten te meten. Bij zulk soort instrumenten hoeft de respondent niet bewust invloed uit te oefenen op de metingen van het instrument. Dit zal in toekomstig onderzoek onderzocht moeten worden.

Bekend is dat naast reclame-inhoud vele andere factoren, zoals emoties, tijdstip van uitzending en betrokkenheid de zapkans zouden kunnen beïnvloeden (Woltman Elpers en Alsem 2001). Alhoewel deze studie rekening houdt met de invloed van commercial en persoonlijke verschillen op de gemiddelde zapkans, zou toekomstig onderzoek moeten uitwijzen hoe deze factoren de zapkans specifiek beïnvloeden in relatie tot “moment-to-moment” reclame-inhoud. Ook moet meer inzicht worden verworven in welke tussenliggende consumentenprocessen de invloed van de reclame-inhoud op het zapgedrag tijdens TV reclames beïnvloeden.

Het in dit artikel gepresenteerde onderzoek geeft een eerste indicatie van de invloed van “moment-to-moment” reclame-inhoud op de zapkans tijdens een specifieke commercial. Reclame-inhoud is ook datgene wat de adverteerder kan beïnvloeden en het (her)ontwerpen van de inhoud is een eerste stap om het aantal zappers tijdens de TV reclame te reduceren.

LITERATUUR

Aaker, David A. Douglas M. Stayman and Michael R. Hagerty (1986), “Warmth in Advertising: Measurement, Impact, and Sequence Effects,” *Journal of Consumer Research*, 12 (March), 365-381.

Baumgartner, Hans, Mita Sujun, and Dan Padgett (1997), “Patterns of Affective Reactions to Advertisements: The Integration of Moment-to-Moment Responses into Overall Judgments,” *Journal of Marketing Research*, 34 (May), 219-232.

Biel, Alexander L. (1998), “Likability: Why Advertising that is Well Liked Sells Well,” in *How Advertising Works: the Role of Research*, John Philip Jones, ed. Thousand Oaks, CA: Sage Publications, 111-120.

Bless, Herbert (2000), "The Interplay of Affect and Cognition: The Mediating Role of General Knowledge Structures," in *Feeling and Thinking: The Role of Affect in Social Thinking*, Joseph P. Forgas, ed. New York: Cambridge Press University, 201-222.

Chaiken, Shelly and Alice H. Eagly (1983), "Communication Modality as a Determinant of Persuasion: The Role of Communicator Salience," *Journal of Personality and Social Psychology*, 45 (August), 241-256.

Fiedler, Klaus (2000), "Toward an Integrative Account of Affect and Cognition Phenomena Using the BIAS Computer Algorithm," in *Feeling and Thinking: The Role of Affect and Social Cognition*, Joseph P. Forgas, ed. New York: Cambridge University Press, 223-252.

Frederick, Shane and George Loewenstein (1999), "Hedonic Adaptation," in *Well-Being: The Foundations of Hedonic Psychology*, Daniel Kahneman, Ed Diener and Norbert Schwartz, eds. New York: Russell Sage Foundation, 302-329.

Jacoby, Jacob (1984), "Perspectives on Information Overload," *Journal of Consumer Research*, 10 (March), 432-435.

Mackie, Diane M. and Leila T. Worth (1991), "Feeling Good, But Not Thinking Straight: The Impact of Positive Mood on Persuasion," in *Emotion and Social Judgments*, Joseph P. Forgas, ed. Oxford: Pergamon Press, 201-219.

Meurs, Lex van, 1999. "Switching during Commercial Breaks". Hilversum: Texom.

Olney, Thomas J., Morris B. Holbrook, and Rajeev Batra (1991), "Consumer Responses to Advertising: The Effects of Ad content, Emotions and Attitude toward the Ad on Viewing Time," *Journal of Consumer Research*, 17 (March), 440-453.

Pasadeos, Yorgo (1990), "Perceived Informativeness of and Irritation with Local Advertising," *Journalism Quarterly*, 67 (1), 35-39.

Pieters, Rik, Luk Warlop and Michel Wedel (2001), "Breaking through The Clutter: Benefits of Advertisement Originality and Familiarity for Attention and Memory," Internal Report, Department of Marketing, Tilburg University, the Netherlands.

Puto, Christopher P. and William D. Wells (1984), "Informational and Transformational Advertising: The Differential Effects of Time," *Advances in Consumer Research*, 11, 638-643.

Ramsey, J. O. and Bernard W. Silverman (1997), *Functional Data Analysis*. New York: Springer.

Resnik, Alan and Bruce L. Stern (1977), "An Analysis of Information Content in Television Advertising," *Journal of Marketing*, 41 (1), 50-53.

Schwartz, Norbert (2001), "Feelings as Information: Implications for Affective Influences on Information Processing," in *Theories of Mood and Cognition: A User's Guidebook*, Leonard L. Martin and Gerald L. Clore, eds. Mahwah, NJ: Lawrence Erlbaum Associates, 159-176.

Siddarth, S. and Amitava Chattopadhyay (1998), "To Zap or Not to Zap: A Study of the Determinants of Channel Switching During Commercials," *Marketing Science*, 17 (2), 124-138.

Trussell, James and Tooni Richards (1985), "Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman-Singer Procedure," in *Social Methodology*, Nancy B. Tuma, ed. San Francisco: Jossey-Bass Inc. Publishers.

Vanden Abeele, Piet and Douglas L. MacLachlan (1994), "Process Tracing of Emotional Responses to TV Ads: Revisiting the Warmth Monitor," *Journal of Consumer Research*, 20 (March), 586-600.

Vilcassim, Naufel J. and Dipak C. Jain (1991), "Modeling Purchase -Timing and Brand-Switching Behavior Incorporating Explanatory Variables and Unobserved Heterogeneity," *Journal of Marketing Research*, 28 (February), 29-41.

Woltman Elpers, Josephine L.C. M. en Karel Jan Alsem, Het voorkomen van zapedrag bij televisiecommercial, *ESB*, 16 februari, 2001, 86e jaargang, 148-151

Woltman Elpers, Josephine L.C. M., Michel Wedel and Rik G. M. Pieters, 2001, The Influence of "Moment-to-Moment" Pleasantness and Informativeness on Zapping TV reclames: A Functional Data and Survival Analysis, Internal Report, Department of Marketing and Marketing Research, University of Groningen, the Netherlands.

8. *Wat onthoudt een consument van een tv-commercial?*

Een kijkje in het brein met neuro-imaging technieken

A. SMIDTS en J.R. ROSSITER

SAMENVATTING

Met een nieuwe neuro-imaging techniek om de activiteit in de hersenen te meten, de zogenaamde steady-state probe topography (SSPT), kan opgespoord worden welke scènes uit een tv-commercial door consumenten goed herinnerd worden. Uit een experiment blijkt dat scènes die langer dan 1,5 seconde duren en scènes die een snelle activatie in de linkerhersen helft veroorzaken, een week later beter herkend worden. Door het toepassen van SSPT kunnen commercials gepretest worden om de herkenning en herinnering zo hoog mogelijk te maken.

1. INLEIDING

Herkenning en herinnering van advertenties en scènes uit commercials zijn veel gebruikte en belangrijke maatstaven om de effectiviteit van reclame vast te stellen. Het onderliggende idee is dat als consumenten op het moment dat zij moeten kiezen zich de merknaam of de essentie van de reclameboodschap herinneren, dit leidt tot een beïnvloeding van hun beslissingsproces. Daarom wordt vaak geprobeerd ‘memorable’ reclame te creëren. Visuele herkenning en herinnering van onderdelen van een tv-commercial wordt derhalve als een bewijs gezien dat de reclame ‘werkt’ en wordt noodzakelijk geacht voor beïnvloeding van het keuzeprocess (Rossiter en Percy, 1997, hoofdstuk 20; du Plessis, 2001).

Bij printreclame is het relatief eenvoudig om te onderzoeken welke delen van een advertentie veel aandacht krijgen en er is al veel bekend over de factoren die daarbij een rol spelen (zie bijvoorbeeld Rossiter en Percy, 1997, hoofdstuk 10). Oogbewegingsonderzoek (eye-tracking) heeft hierbij zijn waarde reeds bewezen (zie bijvoorbeeld Pieters, Rosbergen en Wedel, 1999) en de toepassing van deze techniek vergroot onze kennis en inzicht op dit gebied nog steeds. De techniek om oogbewegingen te registreren en te analyseren is nu commercieel beschikbaar (Verify Nederland) en relatief eenvoudig toe te passen.

Tot voor kort was het niet mogelijk om precies te voorspellen welke scènes uit een tv-commercial nu gemakkelijk onthouden worden. Uiteraard hoopt men dat in elk geval de merknaam en de kern van de boodschap in het lange-termijngeheugen worden opge-

slagen (encoding) en dat deze ook gemakkelijk teruggevonden kunnen worden (retrieval). Met een nieuwe techniek om hersenactiviteit te meten, de zogenaamde SSPT-techniek (steady-state probe topography) is het mogelijk geworden om de effecten van dynamische visuele stimuli (zoals een tv-commercial) te meten. SSPT is een innovatieve versie van het elektro-encefalogram (EEG). Het heeft een hoge temporele resolutie zodat de effecten van snelle beeldwisselingen goed in kaart gebracht kunnen worden. Met SSPT kunnen op een respondentvriendelijke manier 13 opnamen per seconde gemaakt worden van 64 punten op het hersenoppervlak. De techniek is gedurende de laatste twintig jaar ontwikkeld en uitgetest door Richard Silberstein en collega's van het Brain Sciences Institute van de Swinburne University in Melbourne (Australië). De techniek wordt sinds kort door hen commercieel op de markt gebracht onder de naam van NeuroVu (neurovu@neurovu.com). Reclameonderzoek is één van de mogelijke toepassingen.

In dit artikel wordt verslag gedaan van een recent experiment waarin de methode wordt toegelicht en waarin de effectiviteit van de methode om memorabele scènes op te sporen in een tv-commercial wordt aangetoond. Voor details van de techniek en het experiment wordt verwezen naar de recente artikelen van Rossiter, Silberstein, Harris en Nield (2001) en naar Silberstein, Harris, Nield en Pipingas (2000).

De SSPT-techniek wordt allereerst geplaatst in het kader van recente andere technieken van neuro-imaging waarmee hersenstructuren en de hersenwerking in kaart gebracht kunnen worden. Neuro-imaging wordt ook wel brain-imaging of brain mapping genoemd. Deze neuro-imaging technieken worden door neurologen, neuropsychologen, neurobiologen, biofysici en experimenteel psychologen toegepast om de structuur en werking van de hersenen bloot te leggen. Er wordt daarmee gepoogd een dieper inzicht te krijgen in zaken als perceptie, geheugen, taal, aandacht, emotie, etc. Het interdisciplinaire vakgebied wordt wel de cognitieve neurowetenschap genoemd (Hagoort, 2000).

2. METINGEN AAN HET BREIN: NEURO-IMAGING TECHNIKEN

We beperken deze bespreking tot de zogenaamde non-invasieve neuro-imaging technieken waarbij de meting volledig aan de buitenkant van een levende proefpersoon verricht worden. De technieken om hersenactiviteit te onderzoeken worden onderverdeeld in methoden gebaseerd op het meten van *elektrische activiteit* en methoden gebaseerd op het meten van de *doorbloeding* en het gebruik van zuurstof in de hersenen (de zogenaamde hemodynamische methoden).

Elektrische activiteit kan gemeten worden met het elektro-encefalogram (EEG), een techniek die al bestaat sinds 1929, en het magneto-encefalogram (MEG). Beide technieken zijn veilig voor de respondent en hebben een temporele resolutie van enkele milliseconden. Beide zijn daardoor geschikt om de elektrische activiteit continu te meten ('brain waves').

In het bijzonder is men geïnteresseerd in kortstondige veranderingen in het EEG of het MEG die het gevolg zijn van informatieverwerking in de hersenen. Door het meten van ERP's bij EEG (event-related potentials) en event-related fields (ERF's) bij MEG kunnen de effecten van een stimulus vastgelegd worden (Savoy, 2001).

ERP's zijn potentiaalveranderingen die het gevolg zijn van een bepaalde stimulus of van een bepaalde taak. Ze worden gekoppeld aan bepaalde gebeurtenissen (functies) in de hersenen. Voorbeelden van functies zijn: het opslaan van informatie in en het ophalen van informatie uit het lange-termijngeheugen, het richten van de aandacht op een bepaalde positie in de ruimte, etc. De bij dergelijke processen horende elektrische activiteit is echter klein vergeleken met de totale activiteit van de hersenen. Om vast te stellen welk deel van het EEG het gevolg was van de stimulus (en dus een ERP is) wordt de stimulus of taak vaak herhaald, waarna de ERP's bepaald worden door het gemiddelde te nemen van de gemeten signalen (Hendriks, Taatgen en Andringa, 1997). ERP's worden tegenwoordig gemeten met behulp van een groot aantal elektrodes die in een bepaalde configuratie op de schedel zijn bevestigd, of zelfs door deze elektroden in een soort hoofdkapje ('kalotje') te integreren. De techniek is daarom relatief eenvoudig en respondentvriendelijk toe te passen.

MEG is gebaseerd op het meten van zeer kleine magnetische veranderingen in de hersenen als gevolg van hersenactiviteit. Daarvoor zijn Supergeleidende Quantum Interferentie Devices nodig (de zogenaamde SQUIDS). MEG vereist derhalve zeer dure apparatuur en zal daarom minder snel beschikbaar komen voor (commercieel) psychologisch onderzoek. Het is bovendien nog onduidelijk of MEG inderdaad informatie toevoegt aan de informatie uit het EEG of dat het juist vergelijkbare signalen meet. Een nadeel van beide technieken is dat een slecht driedimensionaal beeld van de hersenen wordt verkregen (de spatiale resolutie is vrij laag). Daardoor zijn hersenstructuren niet goed in kaart te brengen met deze technieken. De technieken gebaseerd op veranderingen in doorbloeding (PET en fMRI) zijn daarin beter.

De belangrijkste hemodynamische methoden zijn PET (Positron Emissie Topografie) en fMRI (functionele Magnetische Resonantie Imaging). De achterliggende gedachte bij deze technieken is dat het lokale doorbloedingpatroon in het brein gerelateerd is aan hersenactiviteit. De bloedtoevoer, en dus het zuurstofgebruik, nemen toe bij verhoogde activiteit.

Bij PET wordt een zeer licht radioactieve stof in de bloedbaan ingebracht, via een injectie in de ader van de arm. Het is natuurlijk van belang dat de radioactiviteit terecht komt in de actieve delen van de hersenen. Daarom wordt een stof gebruikt die in de hersenen een bepaalde functie vervult, bijvoorbeeld, zuurstof of glucose. Meestal wordt een radioactieve isotoop van zuurstof (O^{15}) gebruikt die in ongeveer 10 minuten geheel vervallen is tot een niet-radioactieve vorm (Hagoort, 2000). Bij het radioactieve verval wordt een positron uitgezonden die vrijwel meteen botst met een negatief geladen elektron waarbij twee fotonen (gammastralen) ontstaan die in precies tegenovergestelde richting worden uitgezonden. Deze gammastralen worden vervolgens gemeten en zo kan de positie van de oorsprong in het brein zeer nauwkeurig bepaald worden. De proefpersoon ligt daarbij met het hoofd in een cilindervormige detector. Op die manier is op zo'n 5 à 10 mm nauwkeurig vast te stellen waar de activiteit zich bevindt in de hersenen. Door de scan van doorbloeding voor de stimulus te vergelijken met de scan na de invloed van de stimulus (subtrahiemethode) kan vastgesteld worden waar de activiteit zich bevindt. De temporele resolutie is echter zeer laag. Er kan doorgaans slechts 1 PET-scan per minuut gemaakt worden, waarbij alle processen die zich in die minuut afspelen bij elkaar opgeteld worden. Een proefpersoon moet derhalve dezelfde taak vrij lang uitvoeren om inzicht te krijgen. Voor cognitief psychologisch onderzoek is de PET-scan derhalve minder geschikt en tegenwoordig wordt vaker fMRI toegepast.

Bovendien zijn er grenzen aan het aantal keren dat een proefpersoon met radioactief materiaal ingespoten mag worden.

Functionele MRI (fMRI) is een recent alternatief voor PET. Een MRI scanner bestaat uit een supergeleidende magneet, een kleiner aantal elektrische magneten (de zogenaamde gradiëntspoelen) en een reeks computers en versterkers. De proefpersoon ligt in een grote buis. Het magneetveld in de buis is zo sterk (minimaal 1.5 Tesla) dat bepaalde atomen in het lichaam erop reageren. MRI is vooral gebruikt om hersenstructuren in kaart te brengen met een zeer hoge resolutie. fMRI is een nieuwe manier om de MRI apparatuur te benutten om hersen*functies* op te sporen. Bij functionele MRI wordt gebruik gemaakt van het gedrag van hemoglobine in het magneetveld. Hemoglobine is de transporteur van zuurstof in het bloed, en draagt een ijzerion in zich. Deze ijzerkern heeft vrijwel geen effect op het magneetveld als er zuurstof aan het hemoglobine gebonden is. Echter, wanneer de zuurstof is afgestaan verstoort het ijzerion het veld zodanig dat het MR-sigitaal op die plaats zwakker wordt. De mate van zuurstofverzadiging van het bloed verandert bij toegenomen hersenactiviteit. Op deze wijze is precies na te gaan waar in het brein activiteitstoename heeft plaatsgevonden (Hagoort, 2000). Daarbij wordt een beter oplossend vermogen bereikt dan met PET. In de tijd kan het verloop van de hemodynamische respons goed gevolgd worden; het instrument heeft een hoge temporele resolutie.

fMRI is op dit moment één van de meest gebruikte en veelbelovende technieken in de cognitieve neurowetenschap om de werking van hersenen in kaart te brengen (Savoy, 2001). Daarbij wordt het magneetveld nog steeds sterker gemaakt (wel tot 3 tot 4 Tesla of zelfs hoger).

Voor commerciële toepassingen is fMRI nog niet geschikt. Niet alleen is de apparatuur duur, ook de belasting voor de proefpersoon is te groot voor commerciële doeleinden. Zo ligt de proefpersoon in een sterke magneetbuis, wat op zich wel veilig is (zo wordt men bijvoorbeeld niet ingespoten met radioactief materiaal), maar het magneetveld zorgt wel voor interne verhitting zodat een algemene goede gezondheid belangrijk is. De apparatuur veroorzaakt verder een grote hoeveelheid lawaai (de proefpersoon draagt noodzakelijkerwijs gehoorbeschermers); experimenten met audio (of tv-commercials) zijn daardoor uiteraard niet mogelijk.

Een algemene trend en uitdaging in de neurowetenschap is het combineren van de neuro-imaging technieken. De lage spatiële resolutie van EEG en MEG kan goed gecombineerd worden met PET of fMRI data, terwijl de lagere temporele resolutie van fMRI gecompenseerd kan worden met de resultaten van real-time neurale data uit EEG en MEG studies. In combinatie met de kennis uit de experimentele psychologie is dit een veelbelovend onderzoeksveld dat uiteindelijk als doel heeft hersenfuncties, keuzeprocessen en gedrag beter te begrijpen (Savoy, 2001). Het lijkt onontkoombaar dat dit type onderzoek uiteindelijk consequenties zal hebben voor theorieën over consumentengedrag. Tevens kunnen de methoden nu of op termijn gebruikt worden in het marktonderzoek. In de volgende paragraaf wordt specifiek ingegaan op de SSPT-techniek als middel om tv-commercials te analyseren.

3. MEMORABLE MOMENTEN: EEN EXPERIMENT

3.1. Opzet en uitvoering van het experiment

De opzet van het experiment betrof een zogenaamd ‘within-subject’ design: vrouwelijke deelnemers keken naar een 18 minuten durend tv-programma waarin twee blokken van vijf commercials worden uitgezonden. Een week later werd bij dezelfde groep proefpersonen de visuele herkenning van scènes uit de commercials gemeten.

In totaal 35 vrouwen tussen de 25 en 45 jaar werden volgens toeval geselecteerd door een marktonderzoekbureau. Bovendien werden alleen deelnemers in het experiment opgenomen die rechtshandig waren. Geslacht en rechtshandigheid zijn belangrijk om een homogene groep te krijgen waarbij de hersenactiviteit vergeleken kan worden. Bij 97% van de rechtshandige personen is de linkerhersen helft dominant in de verwerking van taal.

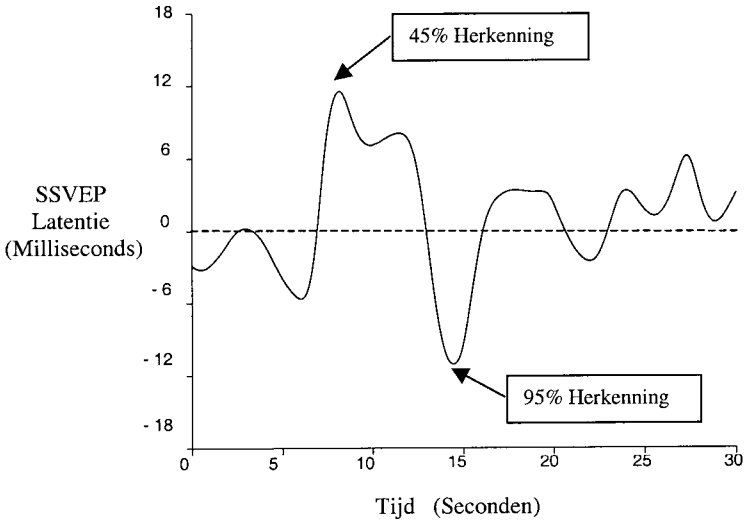
Groepjes van 10 tot 12 proefpersonen nemen plaats in de zaal waarin het tv-programma wordt vertoond. Iedere proefpersoon krijgt een lichtgewicht kapje op het hoofd waarin elektroden gemonteerd zijn die op maximaal 64 plaatsen van de schedel de elektrische activiteit registreren. De signalen worden draadloos doorgestuurd naar de computer. De proefpersonen kunnen dus met volledige bewegingsvrijheid en op een natuurlijke manier naar een tv-scherm kijken. Bovendien krijgt men een bril opgezet met een wit flikkerlicht (het licht flakkert met een frequentie van 13 Hz). Hiermee wordt de ‘baseline’ van de elektrische activiteit in de cortex geïnduceerd. De proefpersoon kan dit flikkerlicht niet waarnemen en merkt daar verder niets van.

De meting van de hersenreactie op het tv-beeld is de latentie (latency) van de SSVEP (steady-state visually evoked potential). Door de latentie te meten kan vastgesteld worden hoe snel of langzaam de hersenen reageren op een inkomende stimulus (een korte latentie is een snelle reactie). Het gaat hierbij om snelheidsverschillen ten opzichte van de baseline. SSVEP latentie veranderingen hangen samen met variaties in de snelheid van neurale informatieverwerking. De veronderstelling is nu dat de latentie samenhangt met de sterkte van geheugenopslag (memory encoding). Beelden waar de hersenen *snel* op reageren worden *beter herinnerd* dan beelden waar de hersenen *langzamer* op reageren. In totaal acht gebieden van de voorhersenen (cerebrale cortex) waarvan bekend is dat deze betrokken zijn bij het verwerken van visuele informatie in het lange-termijngeheugen (Silberstein et al. 1990), werden geselecteerd. De data van twee van die gebieden worden gebruikt in dit experiment.

Er worden 13 metingen per seconde gedaan. Dit betekent dat er bij een 30 seconden durende tv-commercial in totaal 390 metingen beschikbaar zijn. De data wordt vervolgens per proefpersoon uitgebreid bewerkt door een Fourier analyse. Details van deze (technisch ingewikkelde) procedure staan beschreven in Silberstein (1995). Vervolgens wordt de data gemiddeld over alle proefpersonen. Men werkt dus met groepsdata.

Het eindresultaat van de dataverwerking staat gepresenteerd in Figuur 1. Hierin is voor een 30 seconden durende commercial aangegeven hoe de hersenen er op gereageerd hebben. De *minima* van de latenties geven *snelle* hersenreacties aan, de *maxima* de *langzaamste* responses. Zo rond de zevende seconde hebben de hersenen dus erg traag

gereageerd op de commercial, terwijl rond de veertiende seconde een erg snelle reactie optrad. Het maximum in Figuur 1 geeft een langzame reactie aan op dit onderdeel van deze specifieke commercial (deze scène werd een week later door 45% van de respondenten herkend). Het minimum geeft een moment aan dat de hersenen snel reageren (deze scène werd een week later door 95% van de respondenten herkend).



Figuur 1. Een voorbeeld van een SSVEP latencie variatie in de linker hersenhelft tijdens een 30-seconden durende tv-commercial (groepsdata). De horizontale lijn geeft de gemiddelde SSVEP latencie weer van alle commercials samen.

De 35 proefpersonen hebben twee sessies meegemaakt met 1 week tijd tussen de sessies. Bij de tweede sessie werd hen verteld dat zij uitgenodigd werden om vragen te beantwoorden over het tv-programma dat zij de eerste keer gezien hadden. In de eerste sessie werd de hersenactiviteit gemeten tijdens een 18 minuten durende televisiedocumentaire. Deze documentaire werd onderbroken met twaalf tv-commercials van 30 seconden elk. Commercial 1 werd voor het begin en commercial 12 werd na afloop van de documentaire vertoond. Deze twee commercials werden verder niet gebruikt in de test. Commercials 2 t/m 6 en 7 t/m 11 werden in twee blokken van vijf commercials respectievelijk na 4 minuten en na 10 minuten in de documentaire vertoond. De volgorde van deze twee blokken werd voor de helft van de proefpersonen omgedraaid om zodoende te controleren voor volgorde-effecten. Iedere commercial kwam maar 1 keer voor in het experiment.

De commercials waren nieuwe (nog niet vertoonde) commercials van bekende merken. Dit kon gerealiseerd worden door Amerikaanse commercials te gebruiken die in Australië nog niet vertoond waren. Het experiment vond plaats in Australië. De merken waren uiteraard wel in Australië te koop.

Zeven dagen later kregen de proefpersonen in de tweede sessie stilstaande scènes te zien uit de commercial. Proefpersonen dachten een herinneringstest te krijgen over het

tv-programma; iedere suggestie dat het ging om de commercials werd vermeden. In totaal 40 scènes werden geselecteerd uit de 10 commercials: 20 beelden die (op groepsniveau) de snelste hersenreactie hadden opgeroepen en 20 beelden die de langzaamste reactie hadden opgeroepen. Om te voorkomen dat de test beïnvloed werd door bestaande verschillen in merkbekendheid, werden scènes waarin de merknaam te zien was, vervangen door het dichtstbijzijnde beeld zonder de merknaam. Alle 10 commercials bleken vertegenwoordigd te zijn in de selectie van de 40 scènes, alhoewel bepaalde commercials uiteraard meer gedenkwaardige momenten opleveren dan andere!

Een extra set van 40 frames uit andere, eveneens nog niet Australië uitgezonden Amerikaanse commercials, werd geselecteerd als afleidingsmateriaal voor de herkenningstest. Deze 40 scènes kunnen tevens gebruikt worden om de herkenning van de ‘geziene’ scènes te kunnen vergelijken met ‘ongeziene’ scènes. Deze afleidingsframes werden at random tussen de 40 test frames gezet. In totaal kregen de proefpersonen dus 80 beelden te zien die per individu maximaal 5 seconden vertoond werden, en met 1 seconde tussen de beelden. Herkenning werd gemeten door proefpersonen te vragen op de ‘ja’ knop te drukken als zij van mening waren dat de scène vorige week vertoond was, en op de ‘nee’ knop te drukken als deze scène volgens hen vorige week niet vertoond was.

3.2. Resultaten

In totaal 80% van de nieuwe, niet-vertoonde scènes, werd correct geclassificeerd als ‘niet-gezien’. Van de scènes die wél eerder vertoond waren, werd 53% ‘herkend’ en 47% werd dus niet herkend. Deze 53% verschilt niet significant van ‘gokken’ (50% zou dan ‘herkend’ zijn).

Nog even afgezien van de reactie van de hersenen op de commercial, werd wel verwacht dat scènes die langer te zien zijn, ook gemakkelijker worden onthouden en herkend. Daarom werd elektronisch de lengte van iedere scène in de commercial precies opgemeten. Als we de 40 scènes vervolgens splitsen op de mediaan van 1.42 seconden, dan blijkt dat de (20) langere scènes (langer dan 1.42 seconde) inderdaad significant beter herinnerd worden dan de kortere scènes. In totaal 63% van de langere en 41% van de kortere scènes wordt herkend. De correlatie tussen de lengte van scène en herinnering is $r = .41$. Zoals ook al in eerder onderzoek is aangetoond, is er duidelijk een effect van de lengte van de scène op het visuele lange-termijngeheugen.

Vervolgens is gekeken wat het voorspellend effect is van de gemeten hersenactiviteit. Zoals gezegd werd verwacht dat scènes met een snelle reactie van de hersenen beter herkend werden dan die met een langzame reactie. Inderdaad bleek dat het geval te zijn (zie tabel 1): 59% van de snelle reactiescènes werden correct herkend, terwijl 45% van de langzame scènes werden herkend (gepaarde t-test: $t = 4.9$, $p < .0001$). Alhoewel significant, is het absolute verschil van 14 procentpunten niet zo groot.

Vervolgens werd daarom een splitsing aangebracht tussen de reacties gemeten in de linkerhersen helft en die in de rechterhersen helft (19 van de 40 frames bleken gemeten te zijn in de linkerhersen helft, 21 in de rechterhersen helft). Zoals uit tabel 1 blijkt is er geen verschil in herkenning tussen snelle en langzame reacties voor de rechterhersen helft (beide 54% correcte herkenning). Voor de linkerhersen helft is er echter een groot

verschil: 66% van de snelle frames worden herkend tegen 38% van de langzame frames ($t = 7.1, p < .0001$). Herkenning van de snelle reactiescènes is significant hoger dan de gokkans en die van de langzame scènes is significant lager dan de gokkans. Het absolute verschil van 28 procentpunten in herkenning in de linkerhersen helft is maar liefst twee keer zo groot als het verschil dat gemeten is op basis van de gecombineerde hersenhelften.

Uiteraard is 66% herkenning nog geen 100% herkenning. We kunnen ons echter afvragen of 100% een reëel getal is bij deze taak. Ten eerste gaat het om een eenmalige blootstelling aan de commercials en bedraagt de tijd tussen blootstelling en meting maar liefst een week. Als we bovendien de 80% ‘herkenning’ van de niet-vertoonde frames als een benchmark van herkenning opvatten, dan zit 66% dicht tegen dit ‘maximum’ aan. De reden om 80% als maximum te nemen in plaats van 100% is dat proefpersonen wellicht een conservatieve strategie hanteren bij de herkenningstaak. Dat wil zeggen, zij geven aan een scène te herkennen als zij daar absoluut zeker van zijn. In dat geval worden scènes waarover men niet zeker is, vaker als ‘niet herkend’ of ‘nieuw’ geïnclassificeerd. Daarom zou de 80% herkenning van de werkelijk nieuwe scènes als een maximum in deze herkenningstaak opgevat kunnen worden.

Tabel 1. Herkenning van de scènes: totaal en uitgesplitst naar de linker- en rechterhersen helft.

Correcte herkenning (%) van:	Hersen helft		
	Samen	Links	Rechts
Snelle SSVEP response frames	58.7 (20)	66.0 (8)	53.8 (12)
Langzame SSVEP response frames	45.3 (20)	37.9 (11)	54.2 (9)
Aantal respondenten Tussen haakjes is het aantal frames aangegeven	35	35	35

Het effect in de linkerhersen helft blijkt overigens onafhankelijk te zijn van het eerder beschreven effect van de lengte van de scène. Na het controleren voor het effect van scèneduur, bleef de samenhang tussen reactiesnelheid en herkenning positief voor de linker hersen helft ($r = .48$), terwijl die voor de rechterhersen helft verdween ($r = -.17$, ns).

Samengenomen suggereren deze resultaten dat voor vrouwen met een normale laterali-teit (rechtshandig), de linkerhersen helft verantwoordelijk is voor het vastleggen (encoderen) van visuele dynamische scènes in het lange-termijngeheugen.

Tot slot zij opgemerkt dat nadere analyse van de metingen van alle acht de meetgebieden aantonen dat inderdaad *geheugenprocessen* gemeten worden en niet slechts ‘aandacht’ (Silberstein et al., 2000). Het is waarschijnlijk dat een eerdere studie die is uitgevoerd bij consumenten en met tv-commercials (Rothschild en Hyun, 1990) en waarin

de Alpha-activiteit gemeten werd in het EEG, juist wel visuele aandacht in plaats van geheugen (memory encoding) is gemeten.

4. CONCLUSIE

Deze studie geeft aan dat met neuro-imaging technieken (in het bijzonder SSPT) opgespoord kan worden welke scènes uit een commercial memorabel zijn. Op een respondentvriendelijke manier kunnen accuraat de sterke en zwakke punten van commercials onderzocht worden. Hiervoor is een relatief klein aantal respondenten nodig. Na jaren van testen en verfijning wordt de techniek op dit moment commercieel beschikbaar gesteld. Hiermee is een nieuwe en veelbelovende reclame pre-testing methode toegevoegd aan het bestaande assortiment.

Theoretisch is de studie interessant omdat het de eerste studie is waarin de rol van de linkerhersen helft wordt aangetoond bij het transfereren van dynamisch visueel materiaal van het korte naar het lange-termijngeheugen. Dit was verrassend omdat er tot nu toe vanuit wordt gegaan dat de rechterhersen helft erg belangrijk is voor encoderen van visueel materiaal, omdat veel onderzoek heeft aangetoond dat daar de retrieval van visuele informatie (de herkenning van objecten en gezichten) gelokaliseerd ligt. Verbale stimuli worden daarentegen vooral in de linkerhersen helft vastgelegd.

De resultaten van deze studie lijken echter een bevestiging van het HERA-model (het hemispheric encoding /retrieval asymmetry model van Tulving et al., 1994). In het HERA-model wordt gesteld dat de linkerhersen helft een cruciale rol speelt bij het episodische geheugen (het geheugen voor specifieke gebeurtenissen zoals beelden uit een commercial, in tegenstelling tot het semantische geheugen voor feiten en algemene kennis). De rechterhersen helft is dan verantwoordelijk voor het ophalen (de retrieval) van deze beelden uit het lange-termijngeheugen naar herkenning en herinnering. Deze studie heeft laten zien dat visuele informatie inderdaad in de linkerhersen helft vastgelegd wordt. Blijkbaar is niet alleen het vastleggen van verbale informatie, maar ook dat van visuele informatie vooral een activiteit van de linkerhersen helft.

Een andere verklaring voor de bevinding van de belangrijke rol van de linkerhersen helft in dit onderzoek kan zijn dat er alleen rechtshandige vrouwen in de studie zijn opgenomen. Recent onderzoek (o.a. op basis van PET-scan studies, Taylor et al. 1998) suggereert dat de verwerking van 'emotionele' stimuli door vrouwen en mannen in verschillende hersenhelften plaats vindt. Nader onderzoek onder linkshandige vrouwen en onder mannen is daarom noodzakelijk. Voor rechtshandige, vrouwelijke consumenten lijkt echter aangetoond dat de linkerhersen helft cruciaal is bij het onthouden van scènes uit een commercial.

5. IMPLICATIES

Hoe kunnen de resultaten in deze studie nu gebruikt worden om de 'key frames' in een tv-commercial meer memorabel te maken? Ten eerste is het verstandig om dezelfde scène minimaal 2 seconden op het beeldscherm te houden. Dat wil zeggen, cruciaal geachte elementen van de commercial zoals het basisidee van de commercial, het productgebruik of een presentator, moeten minimaal 2 seconden scherp in beeld zijn.

Tevens geldt dit voor het logo of de merknaam. Fast-cuts (korter dan 1.5 seconde) zijn gewoonweg minder memorabel en dus minder effectief (MacLachlan en Logan, 1993).

Een tweede strategie het pre-testen van 'key visuals', bijvoorbeeld door middel van storyboardboards. Met behulp van SSPT kan men nagaan welke visuals memorabele, snelle reacties oproepen in de linkerhersenhelft. Een snelle reactie in de linkerhersenhelft geeft aan dat de scène wordt vastgelegd. Ook het uittesten met SSPT van een ruwe versie van een commercial voordat deze definitief gesneden wordt, zou kunnen aangeven welke onderdelen memorabel zijn en welke absoluut geen enkele rol spelen in herinnering. Het geeft op die manier aan wat de sterke en wat de zwakke momenten in een commercial zijn.

Daarnaast is nader onderzoek nodig om vast te stellen wat voor typen scènes over het algemeen goed onthouden worden. Zo zijn er aanwijzingen dat bijvoorbeeld close-ups van gezichten gemakkelijk herinnerd worden terwijl abstracte scènes minder goed worden vastgelegd. Ook de rol van geluid (tekst en muziek) in het vastleggen van scènes zou onderzocht moeten worden. Met SSPT kan ook gericht gekeken worden naar de combinatie van het vastleggen van de visuele en de audio-inhoud en op welke wijze die elkaar versterken. Bovendien kan vastgesteld worden welke onderdelen van een commercial in het visuele geheugen en welke in het verbale geheugen worden vastgelegd.

De gepresenteerde techniek lijkt een goede kans te hebben uit te groeien tot een nuttige en waardevolle methode voor het pre-testen van reclame. Uiteraard is verder systematisch onderzoek hard nodig om de validiteit, betrouwbaarheid en inzetbaarheid van de methode nader te specificeren. Op de langere termijn zullen ook de andere neuro-imaging technieken mogelijk een rol gaan spelen in de praktijk van het marktonderzoek. Na jaren van te hoog gestemde verwachtingen en vaak teleurstellende resultaten, lijken nieuwe psychofysiologische methoden, zoals neuro-imaging en oogbewegingsonderzoek, een waardevolle bijdrage te gaan leveren.

LITERATUUR

- Hagoort, P. (2000). *De toekomstige eeuw der cognitieve neurowetenschap*. Inaugurele rede, Katholieke Universiteit Nijmegen.
- Hendriks, P., N. Taatgen en T. Andringa (1997). *Breïnmakers en breïnbrekers: inleiding cognitiewetenschap*. Addison Wesley Longman Nederland.
- MacLachlan, J. en Logan, M. (1993). Camera shot length in TV commercials and their memorability and persuasiveness. *Journal of Advertising Research* 33 (2) 57-61.
- Pieters, R., E. Rosbergen en M. Wedel (1999). Visual attention to repeated print advertising: A Test of scanpath theory. *Journal of Marketing Research* 36 (4) 424-438.
- Plessis, E. du (2001). *Reclame en ons brein*. Alphen a/d Rijn: Samsom.
- Rossiter, J.R. en L. Percy (1997). *Advertising Communications & Promotion Management*. 2nd edn. New York: McGraw-Hill.
- Rossiter, J.R., R.B. Silberstein, P.G. Harris en G.A. Nield (2001). Brain-imaging detection of visual scene encoding in long-term memory for TV commercials. *Journal of Advertising Research*, 14 (March/April) 13-21.
- Rothschild, M.L. en Y.J. Hyun (1990). Predicting memory for components of TV commercials from EEG. *Journal of Consumer Research* 16 (4) 472-478.

- Savoy, R.L. (2001). History and future directions of human brain mapping and functional neuroimaging. *Acta Psychologica* 107, 9-42.
- Silberstein, R.B. (1995). Steady state visually evoked potentials, brain resonances and cognitive processes." In: P.L. Nunez (ed.), *Neocortical Dynamics and Human EEG Rhythms*, pp.272-303. New York: Oxford University Press.
- Silberstein, R.B., P.G. Harris, G.A. Nield en A. Pipingas (2000).Frontal steady-state potential changes predict long-term recognition memory performance. *International Journal of Pyschophysiology* 39 (2000) 79-85.
- Silberstein, R.B., M.A. Schier, A. Pipingas, J. Ciorciari en S. Wood (1990). Steady-state visually evoked potential topography associated with a visual vigilance task. *Brain Topography* 3 (1990) 337-347.
- Taylor, S.F., I. Liberzon, L.M. Fig, L.R. Decker, S. Minoshima en R.A. Koepp (1998).The effect of emotional content on visual recognition memory: A PET study. *NeuroImage* 8 (1998) 188-197.
- Tulving, E., S. Kapur, F.I.M. Craik, M. Moscovitz en S. Houle (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission tomography findings. *Proceedings of the National Academy of Science U.S.A.* 91, pp. 2016-2020.

9. Voedselveiligheid: *een kwestie van vertrouwen?*

W. POORTINGA en P. DEKKER*

SAMENVATTING

Risicopercepties, vertrouwen in informatie en steun voor beleid met betrekking tot voedselveiligheid worden in dit hoofdstuk op twee manieren onderzocht. Op basis van Eurobarometer-enquêtegegevens wordt de Nederlandse publieke opinie landenvergelijkend in kaart gebracht en worden statistisch de relaties onderzocht tussen vertrouwen in de veiligheid van voedsel en in informatie daarover. Met focusgroepen wordt gekeken naar processen van meningsvorming en de achtergronden van het toekennen of weigeren van vertrouwen. De conclusie uit de internationale cijfers dat Nederlanders zich relatief weinig zorgen maken over voedselveiligheid en goed van vertrouwen zijn, wordt bevestigd door een nogal relativerende toon in de focusgroepen. Onzekerheid over organisatiebelangen en het ‘gedogen’ van situaties die in strijd zijn met regels doen het vertrouwen in een instantie afnemen.

1. INLEIDING

Het conflict over de Brent Spar, de legionella-besmetting op de Westfriese Flora, de BSE-crisis en controverses rondom biotechnologie en genetische manipulatie laten zien dat milieu- en gezondheidsrisicos actuele thema's zijn. Bovengenoemde incidenten laten ook zien dat overheden en bedrijven de grootste moeite hebben om met risico's om te gaan. Ogenscheinlijk kleine voorvallen kunnen vergaande politieke, economische en sociale gevolgen hebben. Hoewel risico's een onvermijdbaar onderdeel vormen van een moderne technologische samenleving, lijken ze steeds minder geaccepteerd te worden. Het vertrouwen in instanties die verantwoordelijk worden geacht voor risico's, wordt gemakkelijk aangetast. Voor overheden en bedrijven is het dus van groot belang om meer te weten over de ervaringen van risico's in de bevolking en de rol die deze spelen in consumenten- en politiek gedrag, voor de ontvankelijkheid voor informatie en voor het vertrouwen in instanties die de risico's moeten onderkennen, beperken en beheersbaar maken.

Voedselveiligheid krijgt tegenwoordig veel aandacht. Hoewel ons voedsel over het algemeen veel veiliger is geworden, zou het vertrouwen van consumenten in de voedselveiligheid de laatste jaren flink zijn afgenomen (Frewer et al., 1998). Als oorzaak

* Dit hoofdstuk komt voort uit onderzoek van het Sociaal en Cultureel Planbureau voor het Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer (Poortinga et al., 2000).

voor het verminderde consumentenvertrouwen wordt onder andere de toegenomen complexiteit van de voedselproductie genoemd. Doordat productieketens van voedsel langer, internationaler en technischer worden, wordt de productie van voedsel voor de consument ondoorzichtig. Hierdoor zijn de kwaliteit en veiligheid van voedingsmiddelen voor de consument moeilijk te beoordelen. De gevaren mogen dan feitelijk kleiner zijn dan toen voedselvergiftiging een kwestie was van pure pech of goddelijke straf, de omstandigheid dat eventuele problemen nu kunnen worden herleid op menselijk falen maakt het leven een stuk 'riskanter'. Deze omgang met risico's zou volgens Beck (1986) kenmerkend zijn voor het bewustzijn van moderne mensen en instituties van de moderne maatschappij. Giddens (1990) vindt de preoccupatie met risico's evenzeer kenmerkend voor de huidige maatschappij, zij het minder vanwege de feitelijke toename van grootschalige risico's dan vanwege de veranderde aard en de veranderde waarneming van risico's. Ook over de consequenties voor de samenleving verschillen Beck en Giddens van mening. Terwijl Beck de nadruk legt op het ondergraven van vertrouwen in deskundigen, haalt Giddens de onvermijdelijkheid naar voren om toch ergens op te vertrouwen (vgl. Lupton, 1999, Schnabel 1999). Zonder vertrouwen zou de consument immers geen hap meer door de keel krijgen.

2. VOEDSELVEILIGHEID EN VERTROUWEN

Vertrouwen is thans een populair onderwerp in de sociale wetenschappen. Het wordt onderzocht als smeerolie van de economie, een voorwaarde van democratische participatie en als kern van het 'sociaal kapitaal' van een samenleving (Cook, 2000; Dekker, 1999). Ook op het gebied van communicatie en management van risico's kan vertrouwen sinds het afgelopen decennium op toenemende aandacht rekenen (Renn en Levine, 1991; Slovic, 2000; Earle en Cvetkovich, 1995; Frewer et al., 1996; Peters et al., 1997; Johnson, 1999). Het gaat om een complexe materie. Wanneer is er sprake van vertrouwen? Hoe wordt vertrouwen opgebouwd en afgebroken? Verscheidene auteurs hebben hierbij gewezen op het multidimensionele karakter van vertrouwen. *Competentie* wordt conventioneel gezien als belangrijkste factor voor vertrouwen (voor een literatuuroverzicht, zie: Peters et al., 1997; Johnson, 1999). Een informatiebron is geloofwaardiger wanneer wordt gedacht dat de bron verstand van zaken heeft. Volgens anderen is dit te beperkt. Volgens Sandman (1993) heeft verlies in vertrouwen vooral te maken met gebrekkige zorg ('*care*', Johnson, 1999). Sandman doelt hierbij op een (waargenomen) gebrek aan compassie, openheid en rechtvaardigheid. Earle en Cvetkovich (1995) menen dat vertrouwen niet direct gebaseerd hoeft te zijn op duidelijke feiten, maar voor een belangrijk deel door heuristische bepaald wordt. Zo zou vertrouwen vooral gebaseerd zijn op veronderstelde gedeelde waarden ('*consensual values*', Johnson, 1999). Van personen of organisaties, van wie gedacht wordt dat ze dezelfde prioriteiten stellen, kan aangenomen worden dat ze de 'juiste' beslissingen maken.

In samenhang met het bovenstaande zijn er grote verschillen in het vertrouwen dat consumenten stellen in verschillende informatiebronnen. Voor voedselveiligheid blijken consumenten vooral vertrouwen te hebben in wetenschappers, medici en consumentenorganisaties (Frewer et al., 1996). Van deze informatiebronnen wordt verwacht dat ze kennis van zaken hebben, objectief zijn en begaan zijn met het maatschappelijk welzijn. De overheid is niet de meest geschikte informatiebron (Renn en Levine, 1991;

Frewer et al., 1996; Shepherd en Frewer, 1997). Van de overheid wordt vaak gedacht dat ze verkeerde prioriteiten stelt, door bijvoorbeeld meer te letten op economische belangen dan op het maatschappelijk welzijn. Vertrouwen in vrienden, milieugroeperingen en de voedselindustrie is omstreken. Terwijl Frewer et al. (1996) een relatief hoog wantrouwen in vrienden vonden (van hen wordt verwacht dat ze overdrijven en dat ze informatie verdraaien), bleek uit een recent onderzoek dat vrienden juist veel vertrouwen krijgen (Bickerstaff et al., 2001). De ambivalente houding tegenover milieugroeperingen kan verklaard worden uit het feit dat ze begaan zijn met het maatschappelijk welzijn, maar dat ze daarvoor ook selectief omgaan met informatie (Frewer et al., 1996). Wantrouwen in de voedselindustrie zou terug te voeren zijn op commerciële belangen. Al wordt om dezelfde reden soms juist veel vertrouwen in de voedselindustrie gesteld (Frewer et al., 1996). Deze zou om commerciële redenen betrouwbaar zijn omdat verkeerde of misleidende informatie tot negatieve reacties van de consument kan leiden. Het commerciële belang van voedselproducenten is in ieder geval eenduidig, waardoor ze voor de consument 'berekenbaar' zijn.

3. KWANTITATIEF EN KWALITATIEF RISICO-ONDERZOEK

Voor het bestuderen van opvattingen over voedselveiligheid en vertrouwen staan de sociale wetenschapper kwantitatieve en kwalitatieve methoden ter beschikking. Traditioneel verschilt de voorkeur voor een van beide soorten tussen vakgebieden. Zo maakt men in de culturele antropologie voornamelijk gebruik maakt van kwalitatieve ethnografische methoden (participerende observatie) en worden in de (cognitieve) psychologie vooral kwantitatieve methoden gebruikt. Risico-onderzoek is lange tijd gedomineerd geweest door kwantitatieve onderzoeksmethoden. Dit valt vooral toe te schrijven aan het zogenoemde psychometrische paradigma (Slovic, 2000). Recentelijk wordt echter in toenemende mate gebruik gemaakt van kwalitatieve onderzoeksmethoden (Grove-White et al., 1997; Todd et al., 2001). Beide benaderingen hebben hun voor- en nadelen. Kwantitatief onderzoeksmateriaal, gebaseerd op grote aantallen gestandaardiseerde observaties, is goed bruikbaar om algemene opvattingen van het publiek in kaart te brengen. Indien goed opgezet kunnen resultaten uit dit type onderzoek gegeneraliseerd en gereproduceerd worden. Er kunnen regelmatigigheden in relaties tussen risicopercepties, vertrouwen en beleidsvoorkeuren worden ontdekt en veronderstellingen daaromtrent getoetst. Er is onder een aantal sociale wetenschappers echter ook ontevredenheid over deze 'onpersoonlijke' gestandaardiseerde bestudering van menselijke fenomenen, en wordt gepleit voor benaderingen die meer recht doen aan de complexiteit en contextualiteit van risicobeleving. Hiervoor worden door kwalitatief georiënteerde sociale wetenschappers methoden als discours analyse, participerende observatie, en focusgroepen aanbevolen (Flick, 1998). Deze methoden zijn geschikter om te achterhalen hoe personen tot meningen komen en daarbij hun vaak tegenstrijdige overtuigingen combineren. Het ontbreekt ze echter weer aan belangrijke eigenschappen als representativiteit en reproduceerbaarheid.

Met name in de toegepaste sociale wetenschappen worden daarom steeds vaker beide onderzoeksmethoden gecombineerd (Todd et al., 2001; zie voor een voorbeeld van gemengd onderzoek van risicopercepties Waterton en Wynne, 1998). Aangenomen wordt dat een combinatie van kwantitatieve en kwalitatieve onderzoeksmethoden het

onderzoek een meerwaarde geeft. Dit kan op verschillende manieren gebeuren. Het ene onderzoek kan in dienst staan van het andere onderzoek. Zo wordt bijvoorbeeld kwalitatief onderzoek vaak gebruikt om centrale begrippen te identificeren met als doel deze toe te passen in kwantitatief onderzoek. Beide methoden kunnen ook zelfstandig van elkaar uitgevoerd worden. Na de analyses wordt bekeken waarin de resultaten van elkaar verschillen of overeenkomen. De belangrijkste reden om kwantitatieve en kwalitatieve methoden te combineren is *'triangulation'*. Triangulation is gebaseerd op de assumptie dat de beide methoden een andere kijk geven op hetzelfde fenomeen. Door beide methoden toe te passen kan gekeken worden hoe robuust de resultaten zijn.

4. ONS ONDERZOEK

Dit hoofdstuk is een nadere beschouwing op een kortlopend onderzoek dat het Sociaal en Cultureel Planbureau deed in opdracht van het Directoraat-Generaal Milieubeheer van het Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer (Poortinga et al., 2000). Doel van dit onderzoek was om met beperkte middelen meer inzicht te krijgen in de relaties tussen vertrouwen, percepties van voedselgerelateerde risico's en het draagvlak voor beleid (met name overheidsbeleid ten aanzien van bestrijdingsmiddelen). Om hiervan een beter beeld te krijgen hebben we secundaire analyses van internationaal enquêtemateriaal (kwantitatief) gecombineerd met focusgroepen (kwalitatief).

Secundaire survey-analyses

Bij enquête-onderzoek is er zelden een eenvoudige relatie tussen wat men kan meten en wat men wil weten. Metingen ("x% van de Nederlanders vindt y") krijgen doorgaans pas betekenis als ze kunnen worden vergeleken met metingen vroeger of elders. Dit pleit voor het gebruik van eerder en elders gestelde vragen in plaats van het bedenken van nieuwe, misschien meer toegespitste, maar onvergelykbare vragen. Zijn de middelen en de tijd beperkt, zoals in het hier gerapporteerde onderzoek, dan verdient een secundaire analyse van beschikbaar vergelijkend materiaal veelal de voorkeur boven het houden van een kleinschalig eenmalig onderzoek. Zeker als de tijd en het geld voor het goed testen van vragen ontbreken, wegen de voordelen van het operationaliseren van de eigen interesses niet op tegen de nadelen van de onvergelykbaarheid. Niet elk marktonderzoeksbureau zal het met ons eens zijn, maar naar onze overtuiging wordt er vaak veel te snel besloten tot nieuwe datavergaring en te weinig gekeken naar wat beschikbaar is.

Een rijke en gemakkelijk toegankelijke bron vormen de Eurobarometers, de regelmatig in opdracht van de Europese Commissie gehouden opinieonderzoeken in de EU-landen (de data zijn verkrijgbaar via het Steinmetz-archief; een goed overzicht biedt: <http://www.za.uni-koeln.de/data/en/eurobarometer>). In de in het voorjaar van 1998 uitgevoerde Eurobarometer 49.0 is extra aandacht besteed aan de veiligheid van voedsel. Voor de analyses is een selectie gemaakt van vijf landen, namelijk: Nederland, België, Duitsland, het Verenigd Koninkrijk en Denemarken. Met deze selectie kunnen vergelijkingen gemaakt worden tussen Nederland en buurlanden die verschillen in vertrouwen in veiligheid van voedsel en in institutioneel vertrouwen (Dekker, 1999).

Focusgroepen

Focusgroepen komen oorspronkelijk uit de sociologie, meer in het bijzonder het onderzoek naar effecten van massamedia in de jaren veertig van de vorige eeuw (Lazarsfeld, Merton). Vanaf de jaren zeventig wordt de methode populair in het commerciële marktonderzoek, in de jaren tachtig ook in de (verkiezings)campagne-advisering (zie b.v. Anker 1998) en in de jaren negentig groeit ook weer de sociaal-wetenschappelijke belangstelling. Het gaat om gespreksgroepen met meestal tussen de vijf en tien deelnemers en een moderator die op weinig directieve wijze de aandacht gefocust probeert te houden op de onderwerpen en vragen stelt. Naast verbale input zal hij of zij ook vaak gebruik maken van andere middelen zoals stellingen op schrift, al of niet gefingeerde krantenberichten of filmscènes (de methode komt niet voor niets uit de massacommunicatie). Om tot een goed gesprek te kunnen komen wordt veelal enige homogeniteit in leeftijd en opleidingsniveau van de groep nagestreefd, maar soms geeft de vraagstelling van een onderzoek kan ook aanleiding geven voor specifieke contrastgroepen. Naar het ideaal van veel kwalitatief onderzoek houdt men nieuwe groepen tot men 'verzadigd' is, dat wil zeggen geen nieuwe informatie meer wordt geleverd. In marktonderzoek zou dit punt bij vier groepen vaak wel bereikt zijn (of wordt dat in de begroting dwingend verondersteld), maar er zijn ook projecten met een enkele of met tientallen focusgroepen. Idealiter gaat het in focusgroepen om het gesprek als zodanig, om overtuigingsstrategieën, wederzijdse beïnvloeding en processen van collectieve meningsvorming. In feite kan het gesprek soms weinig meer zijn dan een groepsinterview, waarin meningen vooral via de moderator uitgewisseld worden. Sommige verslagen van focusgroepen met hun overdaad aan losse citaten wekken zelfs de indruk dat slechts een aantal individuele interviews door elkaar heen werd afgenomen (Wilkinson, 1998: 196). Met een focusgroep kunnen meningsvorming en de levende publieke opinie beter worden benaderd dan met enquêtes of individuele interviews. Maar ook deze setting blijft een kunstmatige en heeft grote risico's, zoals erg dominante gespreksdeelnemers, een overdreven gerichtheid op consensus, die onderwerpen taboeïseert en afwijkende meningen onderdrukt, of het ontstaan van een atmosfeer waarin deelnemers tegen elkaar gaan opbieden in radicaliteit en bijzondere ervaringen (zie verder Delli Carpini en Williams, 1994; Krueger, 1994; Wilkinson, 1998; Barbour en Kitzinger, 1999).

Voor het kwalitatieve deel zijn in opdracht van het Sociaal en Cultureel Planbureau door onderzoeksbureau Ferro in juli 2000 drie focusgroepen georganiseerd, twee in Zwolle en een in Amsterdam, elk met acht deelnemers. In Zwolle zijn voor een groep middelbaar opgeleiden geselecteerd en voor de ander hoger opgeleiden. In Amsterdam is naar een culturele voorhoede gezocht door hoogopgeleiden te selecteren die biologische levensmiddelen kochten en/of lid waren van ideële organisaties, zoals Greenpeace, Milieudefensie of Amnesty International. In de groepen is gezorgd voor een spreiding in leeftijd (29-49 jaar), sekse (evenveel mannen als vrouwen) en gezinssamenstelling (alleenstaanden, gehuwden of samenwonenden, met en zonder kinderen). Alle deelnemers hebben een onkostenvergoeding ontvangen voor hun deelname. Vragen voor de focusgroepen zijn mede ontleend aan de Eurobarometer. De resultaten van beide methoden zijn daarom gedeeltelijk complementair en kunnen gedeeltelijk gebruikt worden om de robuustheid van de resultaten na te gaan.

5. RESULTATEN VAN DE EUROBAROMETER 49.0

Voedselveiligheid

Tabel 1 laat zien dat Nederlanders (NL) zich, net zoals inwoners van het Verenigd Koninkrijk (VK), weinig zorgen maken over de veiligheid van voedselproducten. Nederlanders blijken vooral weinig vertrouwen in kant-en-klare en andere voorverpakte levensmiddelen te hebben. De inwoners van België (B) en Duitsland (D) hebben het minste vertrouwen in voedselveiligheid. Denemarken (DK) neemt een middenpositie in.

Tabel 1. Wat zijn veilige voedingsproducten (% genoemd).

	NL	B	D	VK	DK
Brood en bakkersproducten	97	86	78	95	88
Vers fruit	93	79	65	91	72
Verse groenten	93	76	64	91	76
Verse vis	75	65	42	84	71
Vers vlees	72	40	34	67	54
Verse melk	92	67	67	89	89
Kaas	95	80	70	89	89
Eieren	84	78	53	78	45
Levensmiddelen in blik	82	53	43	85	51
Diepvriesproducten	87	57	50	86	62
Kant-en-klare producten	55	37	40	61	35
Andere voorverpakte producten	76	38	40	67	37

Tabel 2 toont grote verschillen tussen de landen als het gaat om de plaatsen waar volgens de consument veilige voedingsproducten te vinden zijn. In Nederland en het Verenigde Koninkrijk is een meerderheid van de consumenten van mening dat veilige voedselproducten te vinden zijn in supermarkten en bij kleine winkels. Markten, boeren en kleine producenten worden door relatief weinig Nederlanders genoemd als veilige aankoopplaats. In België en Duitsland vinden consumenten voedingsproducten van supermarkten en kleine winkels minder veilig en worden voedingsproducten van boeren en kleine producenten het veiligst gevonden. Ook worden markten relatief vaak genoemd als veilige aankoopplaats.

Tabel 2. Waar zijn de veiligste voedingsproducten te vinden (% genoemd).

	NL	B	D	VK	DK
Supermarkten	83	39	22	78	53
Kleine winkels	42	22	10	17	32
Markten	18	29	28	9	10
Boeren en kleine producenten	29	49	62	24	39
Ergens anders	3	3	2	2	2
Nergens	1	12	12	2	4

Nederlanders lijken in het algemeen onbezorgder over voedsel dan de inwoners van de andere landen. Ze hebben, met name in vergelijking met Duitsers en Belgen, minder 'romantische' opvattingen over voeding. In Duitsland en België heeft men vooral minder vertrouwen in de veiligheid van 'onnatuurlijke' voorverpakte voedingsmiddelen en heeft men veel vertrouwen in traditionele productiemethoden.

Blijkens Tabel 3 vinden Nederlanders, net zoals inwoners van de andere landen, het geheel vrij zijn van hormonen en het geheel vrij zijn van bestrijdingsmiddelen de meest bepalende factoren voor voedselveiligheid. Ook wordt in alle vijf landen het uitvoeren van controles door bevoegde instanties als een belangrijke factor gezien. Opvallend zijn de verschillen met betrekking tot toevoegingen en conserveringsmiddelen. Terwijl men in België, Duitsland, het Verenigd Koninkrijk en Denemarken vaker van mening is dat voedingsmiddelen alleen veilig zijn als deze *geheel vrij* zijn van toevoegingen, is men in Nederland vaker van mening dat voedingsmiddelen veilig zijn wanneer dit alleen *toegestane* toevoegingen bevat. Dit zou wederom te maken kunnen hebben met het feit dat Nederlanders vrij rationele opvattingen hebben over de veiligheid van voedsel vergeleken met de omliggende landen.

Tabel 3. Factoren die de voedselveiligheid bepalen (% genoemd).

	NL	B	D	VK	DK
Geheel vrij van bestrijdingsmiddelen	53	55	58	57	62
Geheel vrij van hormonen	59	64	55	49	64
Wetenschappelijk voorgeschreven beperking van hoeveelheid bestrijdingsmiddelen en hormonen	30	21	18	26	23
Geheel vrij van toevoegingen	22	37	34	34	31
Bevat alleen toegestane toevoegingen	49	27	41	35	52
Geheel vrij van conserveringsmiddelen	15	32	28	20	19
Bevat alleen toegestane conserveringsmiddelen	44	31	38	36	49
Een geschikte verpakking	35	37	28	27	32
Productcontroles uitgevoerd door bevoegde instanties	58	52	51	40	54

Vertrouwen in informatie

De respondenten is gevraagd in welke mate personen of organisaties de waarheid vertellen over de veiligheid van voedselproducten. Vergeleken met de ons omliggende landen heeft Nederland veel vertrouwen in de verschillende personen en organisaties. In Tabel 4 blijkt dat Nederlanders vooral vertrouwen hebben in consumentenorganisa-

Tabel 4. Betrouwbaarheid van verschillende informatiebronnen.^a

	NL		B		D		VK		DK	
	+	-	+	-	+	-	+	-	+	-
Producenten	17	7	10	19	8	24	11	6	8	11
Kleine winkels	35	5	21	10	19	14	19	10	19	5
Marktkooplui	7	24	18	18	23	16	8	30	5	32
Supermarkten	35	6	16	14	9	24	32	7	13	8
Consumentenorganisaties	76	1	42	8	55	8	48	5	64	2
Landelijke instellingen	50	3	14	18	26	12	26	7	37	4
Europese instellingen	41	5	17	17	17	2	19	9	21	10
Aantal instellingen ^b	2,6	0,5	1,4	1,0	1,6	1,1	1,6	0,8	1,7	0,7

^a + = Spreekt geheel de waarheid (%); - = spreekt in het geheel niet de waarheid (%).

^b Aantal instellingen = het gemiddeld aantal instellingen dat helemaal de waarheid spreekt of dat in het geheel niet de waarheid spreekt.

ties en, zij het in mindere mate, in landelijke en Europese instellingen. Informatie afkomstig van marktkooplui en voedselproducenten wordt het minst betrouwbaar gevonden.

Vertrouwen en risicopercepties

In alle landen blijkt er een duidelijke relatie te zijn tussen vertrouwen en risicopercepties. Meer vertrouwen in informatie over de veiligheid van voedsel gaat gepaard met meer vertrouwen in de veiligheid van voedselproducten. Voor de overzichtelijkheid tonen we in Tabel 5 alleen voor Nederland de samenhang tussen vertrouwen in informatie (tabel 4) en vertrouwen in de veiligheid van voedingsproducten (tabel 2).

Tabel 5. De relatie tussen de betrouwbaarheid van verschillende informatiebronnen en de waargenomen veiligheid van voedingsproducten bij verschillende verkooppunten.

Verkooppunt:		Informatiebron: ^a							
		Supermarkt		Kleine winkels		Marktkooplui		Overheid	
		+	-	+	-	+	-	+	-
Supermarkt	Onveilig	3	56	16	30	25	19	10	53
	Veilig	97	44	84	70	75	81	91	47
Kleine winkels	Onveilig	55	65	42	87	47	70	54	75
	Veilig	45	36	58	13	53	30	46	25
Markten	Onveilig	82	86	79	90	67	96	81	87
	Veilig	18	14	21	10	33	4	19	13
Boeren en kleine producenten	Onveilig	75	57	69	70	54	71	71	60
	Veilig	25	43	31	30	46	29	29	41

^a + = Spreekt geheel de waarheid (%); - = spreekt in het geheel niet de waarheid (%).

De tabel laat ten eerste voor supermarkten, kleine winkels en markten zien dat wie de informatie van een instantie vertrouwt ook meer fiducia heeft in de aldaar verkochte voedingsproducten. Opvallender is dat het vertrouwen in de ene informatiebron vaak ook samengaat met de waargenomen veiligheid van voedingsproducten bij andere verkooppunten. Personen die van mening zijn dat supermarkten geheel de waarheid spreken vinden vaker dat voedingsproducten veilig zijn bij kleine winkels en markten, dan personen die van mening zijn dat supermarkten in het geheel niet de waarheid spreken. Het vertrouwen in informatie van supermarkten gaat echter gepaard met een lager vertrouwen in de veiligheid van voedselproducten van boeren en kleine producenten. Van de personen die van mening zijn dat kleine winkels geheel de waarheid spreken vindt een groter deel dat de voedingsproducten veilig zijn bij supermarkten en markten, maar de voedselveiligheid bij boeren en kleine producenten wordt niet gunstiger beoordeeld. Van de personen die van mening zijn dat marktkooplui geheel de waarheid vertellen, heeft juist een iets kleiner deel vertrouwen in de voedselveiligheid bij supermarkten. In tabel 5 is ook het vertrouwen in informatie van de overheid opgenomen. Met opnieuw een uitzondering voor boeren en kleine producenten, wordt de veiligheid van verkooppunten gunstiger beoordeeld als men meent dat de overheid geheel de waarheid vertelt. Ook blijkt het zo te zijn (hier niet nader geïllustreerd in een tabel) dat personen

die van mening zijn dat de overheid geheel de waarheid spreekt, een groter vertrouwen in de veiligheid van alle in tabel 1 genoemde voedingsproducten. Waarschijnlijk is hier nauwelijks sprake van specifieke causale relaties, maar vooral van een verschil tussen personen: mensen met veel en mensen met weinig vertrouwen.

6. RESULTATEN VAN DE FOCUSGROEPEN

Voedselveiligheid

De deelnemers aan de drie focusgroepen zeggen erg geïnteresseerd te zijn in veiligheid van onze voeding en volgen de berichtgeving op dit gebied. De berichtgeving over voeding brengt enige ongerustheid teweeg. Deze ongerustheid geldt dan vooral de eigen gezondheid en die van partner en kinderen. Ook is een groot deel van de deelnemers geneigd naar aanleiding van verontrustende berichten (een tijdje) meer te letten op wat men koopt, en voeding die negatief in het nieuws is geweest voorlopig te mijden. Veel deelnemers hebben echter de neiging de problemen te relativieren:

Ach, de gemiddelde leeftijd is alleen maar toegenomen. Het eten is hier in ieder geval veiliger dan in India.

Naar eigen zeggen spelen bij de relativering de volgende factoren een rol:

- a) Na aanvankelijke tumult rond een voedselincident vermindert de aandacht van de media over het algemeen weer vrij snel.
- b) In de directe omgeving van de deelnemers is nog nooit iemand ziek geworden als gevolg van voedselincidenten.
- c) Over het algemeen bestaat er relatief veel vertrouwen in controle door overheidsinstanties en het mechanisme van zelfregulering.

Ook in de groeps gesprekken zelf is een patroon van nuanceren en relativering te zien: in alle drie de groepen worden deelnemers die een relatief negatieve kijk op de veiligheid van voeding hebben door de andere deelnemers ‘teruggefloten’ met relativerende opmerkingen ‘dat het allemaal wel meevalt’. Een uitzondering hierop vormt de BSE-affaire. Het feit dat mensen die besmet vlees eten een ernstige en dodelijke ziekte kunnen krijgen, wordt door de deelnemers als zeer bedreigend ervaren.

Vertrouwen in beleid

Over de vraag wie verantwoordelijk is voor de veiligheid van ons voedsel moet men lang nadenken. Dan worden, aarzelend, de Keuringsdienst van Waren of inspectiediensten genoemd als controle instanties. Voor het algemene beleid wordt de overheid of het ministerie ‘dat daar over gaat’ genoemd. Men heeft niet direct een idee welk ministerie dat is. De voedselproducenten en winkeliers worden echter als hoofverantwoordelijke gezien. Wel vindt men vindt het nodig dat ze gecontroleerd worden door onafhankelijke instanties. Daarnaast vindt men dat de consument zelf verantwoordelijk is voor de eigen voedselkeuze. Bovendien kan de consument door middel van zijn koopgedrag voedselproducenten (boeren, levensmiddelenfabrikanten) en detailhandel (supermarkten) bijsturen. In alle groepen wordt opgemerkt dat er een spanningsveld bestaat tussen economische belangen en het maatschappelijk welzijn. In Amsterdam zijn enkele deelnemers de mening toegedaan dat het ‘grote geld’ de doorslag geeft en

het welzijn van de burger daaraan ondergeschikt wordt gemaakt:

Wie het meeste geld heeft, heeft de meeste inbreng. Het welzijn van de burger wordt ondergeschikt gemaakt aan het economisch belang.

Dit roept bij meer positief ingestelde deelnemers tamelijk felle reacties op:

Als je zelfs de overheid niet meer vertrouwt, waar blijf je dan?

De deelnemers hebben het idee dat organisaties zoals de Keuringsdienst van Waren, 'de wetenschap', consumenten-, milieu- en landbouworganisaties sterk hun stempel op het beleid drukken, onder andere door de mobilisatie van burgers (bijvoorbeeld Greenpeace) en door te lobbyen in de politiek (LTO). Onafhankelijke instanties zoals TNO en 'de wetenschap' worden direct door de overheid bij het beleid betrokken.

Over het algemeen zijn de deelnemers de mening toegedaan dat het beleid redelijk effectief is. Men houdt daarbij rekening dat er zich altijd ongelukken kunnen voordoen. De voedselschandalen van de afgelopen jaren beschouwt men als incidenten en niet als een structureel probleem. Het feit dat voedselschandalen in de pers komen, ziet men eerder als een teken dat controle- en signaleringsmechanismen werken dan dat het beleid faalt:

Het feit dat het naar buiten komt is toch een goed teken. Dat betekent dat de controles werken.

In Amsterdam bestaat er bij enkelen wel enig wantrouwen jegens de overheid:

Omdat de overheid geen paniek wil zaaien houden ze dingen achter. En onderzoek duurt altijd verschrikkelijk lang zodat ze veel te laat zijn met waarschuwen.

Anderen in deze groep zijn het echter mee oneens:

Dat wantrouwen deel ik absoluut niet. Ik vind dat harde beschuldigingen. Het feit dat het lang duurt, betekent dat ze het grondig uitzoeken.

Over het algemeen heeft men het gevoel dat de macht van de overheid steeds minder ver reikt. Dit komt vooral door de schaalvergroting in de voedingsindustrie:

Het voedsel komt van steeds verder weg waardoor de condities waaronder het wordt geproduceerd ook steeds minder controleerbaar worden.

Vertrouwen in informatie

In alle drie de groepen is men het eens over wat meer en minder betrouwbare informatiebronnen zijn. De Telegraaf wordt unaniem en zonder enige nuancering als onbetrouwbare informatiebron gezien. Deze krant laat zich volgens de deelnemers in haar berichtgeving vooral leiden door de sensatiezucht van haar lezers. Ook de nieuwsprogramma's van de commerciële omroepen worden minder betrouwbaar geacht. Voor deze omroepen zijn kijkcijfers belangrijker dan objectieve en volledige berichtgeving.

Kranten als de Volkskrant, NRC en Trouw, en nieuws- en actualiteitenprogramma's van de publieke omroepen worden wel betrouwbaar gevonden. Overigens gaan de deelnemers ervan uit dat geen van de genoemde bronnen pertinente leugens verkondigt. Wel kunnen bronnen selectief zijn in het geven van informatie en bepaalde feiten benadrukken of juist weglaten waardoor een verkeerd beeld van het probleem ontstaat. Deze selectiviteit wordt ingegeven door een bepaald belang. Zo zal de overheid er belang bij hebben om paniek te vermijden en om die reden bepaalde feiten verzwijgen. Milieudefensie zal informatie geven vanuit het milieubelang en de Consumentenbond vanuit het belang van de consument. Bij de beoordeling van een informatiebron houdt men ook rekening met eventuele commerciële belangen. Deze kunnen een reden zijn om bepaalde feiten te verzwijgen en andere aan te dikken:

Greenpeace is de laatste jaren steeds commerciëler geworden. Die willen vooral zoveel mogelijk leden werven. Om die reden dikken ze de problemen, zoals toen met die Brent Spar affaire, nog extra aan.

Commerciële belangen kunnen ook tot vertrouwen leiden:

Albert Heijn kan het zich toch niet veroorloven dat er echt iets mis gaat. Die hebben er alle belang bij dat de groenten en fruit in hun winkels veilig zijn.

Enkele gespreksdeelnemers proberen actief informatie uit verschillende bronnen te verzamelen en te vergelijken om tot een inschatting van de ernst van een bepaald probleem te komen. De veelheid van soms tegenstrijdige informatie wordt in het algemeen positief gewaardeerd:

Organisaties houden elkaar in evenwicht door de materie vanuit verschillende perspectieven te belichten.

7. TOT BESLUIT

Doel van het onderzoek was inzicht te krijgen in de relatie tussen vertrouwen, perceptie van voedselgerelateerde risico's en draagvlak voor beleid. Uit de Eurobarometer blijkt dat Nederland een bijzondere plaats inneemt als het gaat om vertrouwen in de veiligheid van voedsel. Nederlanders maken zich over de hele linie weinig zorgen over de veiligheid van voedsel in vergelijking met de ons omliggende landen. In België en Duitsland heeft men over het algemeen het minste vertrouwen in voedselveiligheid. Dit hangt waarschijnlijk samen met de vele voedselincidenten die zich met name in België recentelijk hebben voorgedaan. Ook de opvattingen over veilige verkooppunten verschillen. Nederlanders hebben veel vertrouwen in voedingsproducten van supermarkten en kleine winkels. In België en Duitsland stelt men vertrouwen stelt in voedingsmiddelen van boeren en kleine producenten. Hieruit zou men kunnen concluderen dat Duitsers en Belgen meer 'romantische' opvattingen hebben over voeding. Voedingsmiddelen zijn veilig als ze maar op traditionele wijze geproduceerd worden. Dit zou ook kunnen blijken uit het gegeven dat men minder vertrouwen heeft in de veiligheid van 'onnatuurlijke' voorverpakte voedingsmiddelen.

Naast een groot vertrouwen in voedselveiligheid hebben Nederlanders ook meer vertrouwen in informatie over voedselveiligheid. Weer zijn de verschillen met België en Duitsland het grootst. Nederlanders blijken vooral meer vertrouwen te hebben in controlerende landelijke en Europese instellingen. Ook uit de resultaten ten aanzien van de factoren die van invloed zijn op de voedselveiligheid valt af te lezen dat Nederlanders veel vertrouwen hebben in controlerende en regelgevende instanties: terwijl men in de omringende landen vaker van mening is dat voedingsmiddelen alleen veilig zijn wanneer deze geheel vrij zijn van toevoegingen en conserveringsmiddelen, is men in Nederland vaker van mening dat voedingsmiddelen veilig zijn wanneer deze niet meer dan de toegestane toevoegingen en conserveringsmiddelen bevatten.

Er is een duidelijk verband tussen vertrouwen in informatie en vertrouwen in voedselveiligheid: er is niet alleen een relatie tussen het vertrouwen in een informatiebron en de aldaar te verkrijgen voedingsproducten, maar ook tussen vertrouwen in de overheid en het vertrouwen in de veiligheid van voedingsproducten. Waarschijnlijk speelt hier dat er mensen zijn die over de hele linie meer vertrouwen hebben en mensen die dat over de hele linie minder hebben. Overigens hebben we een dergelijke onderscheiding niet kunnen maken op basis van een hier niet verder gerapporteerde analyse van sociaal-demografische groepen. Waarschijnlijk zijn risicopercepties en vertrouwen meer geworteld in culturele oriëntaties die daar betrekkelijk los van staan. In dit verband lijkt het zinvol te kijken in hoeverre de tegenwoordig in de risicoliteratuur populaire 'Culturele theorie' (Douglas en Wildavsky, 1982; Dake, 1992 meer licht op de verschillen kan werpen.

De resultaten van de focusgroepen komen voor een groot deel overeen met de bevindingen uit de literatuur en het kwantitatieve deel. De deelnemers aan de discussies zijn erg geïnteresseerd in voedselveiligheid, maar maken zich nauwelijks zorgen over dit onderwerp. In de focusgroepen wordt wel met lichte zorg gesproken over het steeds grootschaliger worden van de voedselproductie, waardoor men minder zicht heeft op wat er tijdens productie, verwerking en opslag gebeurt. De consument kan aan de levensmiddelen in de winkel niet zien of proeven dat ze verontreinigd of dat ze besmet zijn met bijvoorbeeld salmonella. Relativering lijkt de meest voorkomende reactie te zijn in de focusgroepgesprekken. Men wijst erop dat er in Nederland zelden doden vallen en dat de gemiddelde leeftijd alleen maar hoger is geworden door betere voeding. Elders in de wereld is het veel slechter gesteld met de veiligheid van voedsel. Er bestaat in het algemeen dan ook veel vertrouwen in het Nederlandse regelgeving en controle. Verder denkt men dat zelfregulering bijdraagt aan veiligere voeding: voedingsproducenten en -leveranciers hebben er (commercieel) belang bij dat er geen ongelukken gebeuren. Enkelen die hierbij hun twijfels hebben kregen nauwelijks bijval. Dat er af en toe iets mis gaat, vindt men onvermijdelijk. Men gaat er vanuit dat de bevoegde instanties dan afdoende maatregelen nemen. Voedselaffaires zijn voor de meeste mensen incidenten en niet het gevolg van een structureel probleem. Het is wel belangrijk dat problemen tijdig gesignaleerd worden en dat dit direct in de openbaarheid wordt gebracht. Men hecht daarom ook veel belang aan 'waakhonden', zoals Greenpeace, Milieudefensie en de Consumentenbond, die regelgevende en controlerende instanties kritisch volgen.

Op het gebied van geloofwaardigheid van verschillende informatiebronnen laten de

gebruikte onderzoeksmethoden interessante verschillen zien. Terwijl de Eurobarometers internationaal ondubbelzinnige en dank zij grote steekproefaantallen ‘significante’ verschillen in betrouwbaarheid tussen informatiebronnen tonen, blijkt het in de focusgroepen een stuk lastiger om een onderscheid te maken tussen betrouwbare en onbetrouwbare informatiebronnen. Men denkt niet dat er bronnen zijn die aperte leugens verkopen. Men gaat er eerder vanuit dat informatiebronnen selectief te werk gaan door bepaalde feiten te benadrukken en andere te verzwijgen. Men is het erover eens dat de publieke omroepen en de ‘kwaliteitskranten’ streven naar objectiviteit. De geringere betrouwbaarheid van de Telegraaf is in de groepen geen punt van discussie, maar eerder een openbare belijdenis. Ook de commerciële omroepen zouden minder objectief zijn omdat zij kijkcijfers laten prevaleren boven waarheidsvinding.

Bij de beoordeling van informatie speelt het begrip ‘belang’ een grote rol. Gedragingen van personen en instanties worden verklaard uit hun mogelijke belangen. Men gaat ervan uit dat elke bron er belang bij heeft om bepaalde feiten te vermelden en om andere te verzwijgen. Zo veronderstelt men dat de overheid geneigd zal zijn om feiten te verzwijgen om paniek te voorkomen, hoewel dit vertrouwen ondergraaft. Een informatiebron kan haar geloofwaardigheid verhogen door feiten te geven die klaarblijkelijk tegen haar belang indruisen.

De combinatie van kwantitatief (Eurobarometer) en kwalitatief (focusgroepen) materiaal bleek een goede methode om een compleet beeld van opvattingen over voedselveiligheid te krijgen. Alleen de rapportage van de Eurobarometer zou een pover beeld hebben opgeleverd. Met de focusgroepen is inhoudelijk dieper gegraven en wordt inzicht geboden in argumentaties die mensen hanteren, al zijn door de aard van het onderzoek en het geringe aantal participanten de argumenten zelf getalsmatig niet te extrapoleren. Hoewel de beide bronnen niet optimaal op elkaar waren afgestemd (de Eurobarometer lag er al), bleken de resultaten verrassend goed op elkaar aan te sluiten.

De algehele conclusie kan zijn dat voedselveiligheid in Nederland geen grote publieke zorg is. Voedselveiligheid is een interessant onderwerp, maar van individueel risicobesef lijkt nauwelijks sprake. Hier lijkt te gelden wat eerder in landenvergelijkend opinieonderzoek over milieuvervuiling werd gevonden: voor de gemiddelde Nederlander is het eerder een dringend maatschappelijk probleem dan iets waar hij of zij persoonlijk last van heeft (Van den Broek et al., 1997). Opgemerkt zij wel dat ons enquêtémateriaal uit 1998 stamt en de discussiegroepen medio 2000 zijn gehouden. Nieuwe incidenten en de discussies over de intensieve veeteelt naar aanleiding van de MKZ-crisis en over genetisch gemodificeerd voedsel kunnen nadien het probleembesef hebben versterkt. Daarnaast blijft het moeilijk uit te maken in hoeverre het onderwerp werkelijk weinig zorgen baart en in hoeverre men toch niet ook de ongerustheid ‘wegpraat’. Voor het laatste zijn in de focusgroepen wel enkele aanwijzingen te vinden. Hoe dat ook zij, er is geen sterke *publieke* opinie voor ingrijpend nieuw beleid om risico’s uit te sluiten. Men heeft veel vertrouwen in overheidscontroles, maar de focusgroepen laten ook zien dat dit vertrouwen wordt ondermijnd door ‘gedoogbeleid’. Over de aanvaardbaarheid van risico’s kun je van mening verschillen, maar eenmaal vastgestelde normen oprekken, uitzonderingen accepteren, overtredingen aanvaarden, daarvoor bestaat in de focusgroepen geen enkel begrip.

LITERATUUR

- Anker, H. (1999). Sterk en sociaal: ontwikkeling van de PvdA-boodschap in de campagne van 1998. In: A.E. Bronner et al. (red.), *Recente ontwikkelingen in het marktonderzoek, jaarboek 1999*. Haarlem: De Vrieseborch.
- Barbour, S. en J. Kitinger (red., 1999). *Developing focus group research*. Londen: Sage.
- Beck, U. (1986). *Risikogesellschaft*. Frankfurt am Main: Suhrkamp.
- Bickerstaff, K., S. Gerrard, I. Langford, T. O'Riordan, N. Pidgeon, J. Niewoehner, W. Poortinga en P. Simmons (2001). *Preliminary Results from a public attitudes and behaviour survey on the foot and mouth disease outbreak*. Norwich: Centre for Environmental Risk [Internet: <http://www.uea.ac.uk/env/cer/fmdresults.htm>].
- Broek, A. van den, P. Dekker en M. Nas (1997). De publieke opinie als draagvlak voor het milieubeleid. In: A.E. Bronner et al., (red.), *Recente ontwikkelingen in het marktonderzoek, jaarboek 1997*. Haarlem: De Vrieseborch.
- Cook, K.S. (red.; 2000). *Trust in society*. New York: Russell Sage Foundation.
- Dake, K. (1992). Myths of nature. *Journal of Social Issues* 48 (4), 21-37.
- Dekker, P. (1999). Niveaus van vertrouwen in landen van de Europese Unie. In: A.E. Bronner et al., (red.), *Recente ontwikkelingen in het marktonderzoek, jaarboek 1999*. Haarlem: De Vrieseborch.
- Delli Carpini, M.X. en B. Williams (1994). The method is the message. In: M.X. Delli Carpini, L. Huddy en R.Y. Shapiro (red.), *Research in micropolitics*. Greenwich: JAI Press.
- Douglas, M. en A. Wildavsky (1982). *Risk and culture*. Berkeley: University of California Press.
- Earle, T.C. en G.T. Cvetkovich (1995) *Social trust*. Westport (CT)/Londen: Praeger.
- EC (1999). *Eurobarometer 49.0*. Brussel: Europese Commissie, DG voor onderwijs en cultuur [Internet: <http://europa.eu.int/comm/dg10/epo/>].
- Fife-Shaw, C. en G. Rowe (1996). Public perceptions of everyday food hazards. *Risk Analysis* 16, 487- 500.
- Flick, U. (1998). *An introduction to qualitative research*. Londen: Sage.
- Franklin, J. (red., 1998). *The politics of risk society*. Cambridge: Polity Press.
- Frewer, L.J., C. Howard, D. Hedderley en R. Shepherd (1996). What determines trust in information about food-related risks? *Risk Analysis* 16, 473-486.
- Frewer, L.J., C. Howard, D. Hedderley en R. Shepherd (1998). Methodological approaches to assessing risk perception associated with food-related risks. *Risk Analysis* 18, 95-102.
- Giddens, A. (1990). *The consequences of modernity*. Cambridge: Polity Press.
- Grove-White, R., P. Macnaghten, S. Mayer en B. Wynne (1997). *Uncertain world*. Lancaster: Lancaster University, CSEC.
- Johnson, B.B. (1999). Exploring the dimensionality in the origins of hazard-related trust. *Journal of Risk Research* 2 (4), 325-354.
- Krueger, R.A. (1994). *Focus groups*. Thousand Oaks: Sage.
- Lupton, D. (1999). *Risk*. Londen/New York: Routledge.
- Margolis, H. (1996). *Dealing with risk*. Chicago: University of Chicago Press.
- Marris, C., I.H. Langford en T. O'Riordan (1998). A quantitative test of the cultural theory of risk perception. *Risk Analysis* 18, 635-647.
- Peters, R.G., V.T. Covello en D.B. McCallum (1997). The determinants of trust and credibility in environmental risk communication. *Risk Analysis* 17, 43-54.
- Poortinga, W., P. Dekker en J. van Male (2000). *Vertrouwen en risicoperceptie*. Den Haag: Sociaal en Cultureel Planbureau [Internet: <http://www.scp.nl/users/dekker/papers/Risico.pdf>].
- Renn, O. en D. Levine (1991). Credibility and trust in risk communication. In: R.E. Kasperson en P.M. Stallen (red), *Communicating risks to the public*. Amsterdam: Kluwer.
- Sandman, P.M. (1993). *Responding to community outrage*. Fairfax: American Industrial Hygiene Association.
- Schnabel, P. (1999). Zekerheid en de noodzaak van vertrouwen. *Economisch Statistische Berichten* 8, 10-12.

- Shepherd, R. en L.J. Frewer (1997). Risk communication. In: D. R. Tennant (red.), *Food chemical risk analysis*. Londen: Chapman and Hall.
- Slovic, P. (2000). *The perception of risk*. Londen: Eartscan.
- Smithson, J. (2000). Using and analyzing focus groups: limitations and possibilities. *International Journal of Social Research Methodology* 3, 2, 103-119.
- Sparks, P. en R. Shepherd (1994). Public perception of the potential hazards associated with food production and food consumption. *Risk Analysis* 14, 799-806.
- Todd, Z., B. Nerlich, B., S. McKeown en D.D. Clarke (2001). *Mixing methods in Psychology*. London: Routledge (in press).
- Waterton, C. en B. Wynne (1998). Can focus groups access community views? In: Barbour en Kitzinger, 127-143.
- Wilkinson, S. (1998). Focus group methodology. *International Journal of Social Research Methodology* 1, 181-203.

10. Steun voor seksdiscriminatie op de arbeidsmarkt in Nederland: individuele en maatschappelijke verklaringen

M. GESTHUIZEN, P. SCHEEPERS en M. VERLOO

SAMENVATTING

In deze bijdrage gaan we na in hoeverre er in de jaren 1979-1994 een neerwaartse trend heeft bestaan in de steun voor seksdiscriminatie op de arbeidsmarkt. Die steun blijkt geleidelijk te zijn afgenomen van 31% in 1979 naar 13% in 1994. Mensen die zich verwant voelen met de cultureel conservatieve ideologie hechten hun goedkeuring aan seksdiscriminatie: mensen die behoren tot een kerkgemeenschap of mensen die veelvuldig naar de kerk gaan, laag opgeleiden, mensen geboren voor de Tweede Wereldoorlog, huisvrouwen en kleine zelfstandigen. Daarnaast blijkt dat periodieke omstandigheden als werkloosheid, economische groei en het aantal vrouwen op de arbeidsmarkt negatieve effecten hebben op de steun voor seksdiscriminatie.

1. INLEIDING

Met zekere regelmaat verschijnen onderzoeken die aan de arbeidsmarkt gerelateerde verschillen tussen mannen en vrouwen beschrijven en trachten te verklaren. Telkens blijkt weer dat vrouwen in een achterstandspositie verkeren ten opzichte van mannen. In het verleden is er altijd een verschil in werkloosheidspercentages ten nadele van vrouwen geweest (Keuzenkamp & Oudhof, 2000). In 1999 was de werkloosheid onder vrouwen ongeveer twee keer zo groot als onder mannen (Keuzenkamp & Oudhof, 2000). Een andere maat waaruit een achterstandspositie blijkt, is het bruto uurloon van vrouwen als percentage van het bruto uurloon van mannen. Dit percentage ligt tussen 1995 en 1997, uitgesplitst naar beroepsniveau, altijd tussen de 73 en 85 procent. Slechts de evenredig geworden verdeling van mannen en vrouwen over de verschillende beroepsniveaus schijnt te wijzen op een convergentie tussen de arbeidsmarktkansen van de seksen. Wanneer we rekening houden met het toegenomen opleidingsniveau en de stijgende arbeidsmarktparticipatie van vrouwen ten opzichte van mannen (van der Lippe & van Doorne-Huiskes, 1995), mogen we spreken van een discrepantie tussen de aanwezigheid van kennis en vaardigheden bij vrouwen, en de feitelijke arbeidsmarktkansen en de beloningen die ze daarvoor terugkrijgen.

Gezien deze discrepantie kan de conclusie getrokken worden dat vrouwen op de arbeidsmarkt nog steeds benadeeld worden ten opzichte van mannen. Daarom is het relevant na te gaan in hoeverre deze situatie draagvlak heeft onder de Nederlandse bevolking. Steun voor seksdiscriminatie op de arbeidsmarkt zal in dit onderzoek geïn-

diceerd worden door situaties voor te leggen waarin hetzij een man hetzij een vrouw ontslagen moet worden respectievelijk een promotie kan krijgen, terwijl de kandidaten voor het overige niet van elkaar verschillen. Gelijkaardige indicatoren zijn gangbaar in onderzoek naar steun voor etnische discriminatie (Sociaal en Cultureel Planbureau, 1994; Coenders & Scheepers, 1998).

We verwachten dat tussen 1979 en 1994, de tijdsspanne die onderzocht zal gaan worden, steeds minder Nederlanders bereid zullen zijn vrouwen op de arbeidsmarkt te benadelen. Deze verwachting is ingegeven door de bevinding dat de Nederlandse bevolking steeds meer afstand heeft genomen van de traditionele rolverdeling tussen mannen en vrouwen (Hooghiemstra & Hiphuis-Nell, 1993). Concreet trachten we een antwoord te vinden op de volgende centrale vraagstellingen. (1) *In hoeverre is er in Nederland tussen 1979 en 1994 sprake van een neerwaartse trend wanneer het gaat om de steun voor seksdiscriminatie op de arbeidsmarkt?* (2) *Welke sociale categorieën steunen seksdiscriminatie op de arbeidsmarkt in Nederland tussen 1979 en 1994?* En (3) *welke periode-, en/of cohortkenmerken kunnen de neerwaartse trend verklaren?*

2. THEORIE EN HYPOTHESEN

Individuele socialisatie

Hoewel er geen directe theoretische noties bestaan op grond waarvan we kunnen verwachten bij wat voor mensen steun voor seksdiscriminatie op de arbeidsmarkt bestaat, beschikken we wel over indirecte noties. We veronderstellen dat steun voor seksdiscriminatie is ingegeven door de opvatting dat vrouwen hun traditionele rol dienen te vervullen. Uit eerdere studies weten we voorts dat deze traditionele opvatting over de rol van vrouwen is ingebed in de zogenaamde cultureel conservatieve ideologie (Middendorp, 1991; Scheepers et al., 1999), een ideologie die impliceert dat men moet handelen volgens traditionele normen en vooral populair is bij kerkleden en oudere cohorten. De verklaring die hiervoor vaak geopperd wordt, is dat kerkleden gesocialiseerd zijn in instituties waarin dergelijke traditionele waarden overgebracht worden. Voor mensen die geboren zijn voor de Tweede Wereldoorlog zou kunnen gelden dat zij zijn gesocialiseerd met deze cultureel conservatieve ideologie, die toen immers nog wijder verbreid was. De eerste algemene propositie luidt dan ook dat naarmate men een meer cultureel conservatieve ideologie onderschrijft, men des te sterker seksdiscriminatie op de arbeidsmarkt steunt. Teneinde van deze algemene propositie ook toetsbare hypothesen af te leiden, dienen we te weten te komen welke sociale categorieën deze cultureel conservatieve ideologie een warm hart toedragen, hetgeen reeds bekend is uit eerder onderzoek (Scheepers, Peters & Felling, 2000). Op grond daarvan komen we tot de volgende toetsbare hypothesen. *In Nederland steunen tussen 1979 en 1994 (H1a) gereformeerden, (H1b) katholieken, (H1c) Nederlands hervormden, (H1d) personen geboren voor 1940 (H1e) personen zonder een diploma, (H1f) een LBO-diploma, (H1g) ongeschoolde handarbeiders en (H1h) huisvrouwen, (H1i) kleine zelfstandigen en (H1j) mannen seksdiscriminatie op de arbeidsmarkt meer dan de gemiddelde Nederlander.*

Maatschappelijke socialisatie

Naast de invloed van de in bovenstaande paragraaf besproken socialisatiefactoren op individueel vlak, verwachten we tevens socialiserende veranderingen aan te treffen op

maatschappelijk niveau. De vraag is of het feit dat steeds meer vrouwen toegetreden zijn tot de arbeidsmarkt tot toegenomen gewenning heeft geleid, zodat men positiever is gaan denken over de deelname van de vrouw op de arbeidsmarkt. Dit zou tot uiting moeten komen in een afnemende steun voor seksdiscriminatie op de arbeidsmarkt. Op grond hiervan formuleren we de tweede algemene propositie van dit onderzoek: hoe meer men gewend raakt aan het feit dat vrouwen zich op de arbeidsmarkt begeven, des te minder zal men seksdiscriminatie op de arbeidsmarkt steunen. Ten einde van deze propositie meer specifieke hypothesen af te leiden, maken we gebruik van de Critical Mass Theory. Moss Kanter (1977) gaat in haar werk vooral in op de gevolgen van structurele kenmerken van grootschalige arbeidsorganisaties op het gedrag van de leden van die organisaties. Sociale categorieën die slechts in beperkte mate vertegenwoordigd zijn, zijn weliswaar zichtbaar, maar zijn meer geïsoleerd en buitengesloten van informele netwerken, lopen kans om gestereotypeerd te worden in tegenstelling tot mensen behorend tot de meerderheid. Wanneer slechts een klein aantal van de minderheidsgroepering aanwezig is, wordt zowel de groepsidentiteit van de meerder- als de minderheid sterk geprofileerd. Wanneer de getalsverhouding echter meer en meer in de richting van evenredigheid groeit, zijn leden van de minderheid in staat de cultuur van “de groep” te beïnvloeden. Volgens Moss Kanter ligt er een omslagpunt tussen wat zij “skewed” en “tilted groups” noemt. Wanneer de minderheidsgroepering een percentage tussen de 15 en 35 procent bereikt, is zij in staat deze invloed uit te oefenen. Sindsdien wordt het ijkpunt van 30 procent beschouwd als dit omslagpunt, ofwel de “critical mass”.

Dahlerup (1988) heeft deze theorie gebruikt om verwachtingen te formuleren omtrent het gedrag van politici als de man-vrouw-verhoudingen zo veranderen dat de categorie vrouwen de “critical mass” overstijgt. Het idee is nu dat een cultuurverandering plaatsvindt wanneer het percentage deelnemende vrouwen op de arbeidsmarkt toeneemt in de zin dat er een bepaalde gewenning optreedt. Hoe hoger dus het percentage vrouwen op de arbeidsmarkt, en meer specifiek, wanneer de “critical mass” de 30 procent bereikt, des te meer raakt men gewend aan het feit dat vrouwen zich op de arbeidsmarkt begeven. Zodoende luidt onze hypothese met betrekking tot het percentage vrouwen op de arbeidsmarkt: *(H2a) hoe hoger het percentage vrouwen op de arbeidsmarkt, des te minder steunt men seksdiscriminatie op de arbeidsmarkt.* Dezelfde redenatie kan gebruikt worden wanneer het gaat over de invloed van de formatieve jaren, ofwel de socialiserende werking van de tijd waarin men opgroeide, op huidige attitudes. Mannheim (1964) veronderstelde dat personen geboren in ongeveer dezelfde periode, dezelfde maatschappelijke omstandigheden ervaren. Deze omstandigheden, aanwezig in een vormende levensfase, zouden ervoor kunnen zorgen dat de houdingen en gedragingen die hieruit ontstaan relatief stabiel blijven gedurende de rest van het leven. Op grond van deze ideeën komen we tot de volgende hypothese: *(H2b) hoe hoger het percentage vrouwen op de arbeidsmarkt gedurende de formatieve levensfase was, des te minder steunt men in het heden seksdiscriminatie op de arbeidsmarkt.*

Macro-economische kenmerken

De beschreven maatschappelijke socialisatie impliceert een graduele daling van de steun voor seksdiscriminatie op de arbeidsmarkt. Het lijkt echter niet plausibel te veronderstellen dat deze dalende trend zuiver lineair verloopt. We verwachten dat men in slechtere economisch tijden geneigd is om traditionele waarden te verkiezen. De Conflict Theorie (vgl. Coenders & Scheepers, 1998) lijkt zich uitstekend te lenen voor het

afleiden van hypothesen over de invloed van maatschappelijke omstandigheden op de steun voor seksdiscriminatie in een bepaalde periode. Eveneens biedt deze theorie houvast om hypothesen over cohort effecten af te leiden.

Het vertrekpunt van de Conflict Theorie is “het axioma van fundamenteel conflicterende belangen” (Coser, 1956; Blalock, 1967). In de samenleving trachten verschillende sociale categorieën, zoals mannen en vrouwen, schaarse hulpbronnen te bemachtigen. Omdat het schaarse karakter van de hulpbronnen er toe leidt dat er conflicterende belangen ontstaan, zullen leden van sociale categorieën andere sociale categorieën als potentiële concurrenten beschouwen.

Uit de Conflict Theorie leiden we af dat de mate van steun voor seksdiscriminatie op de arbeidsmarkt in een bepaalde periode beïnvloed wordt door maatschappelijke omstandigheden in deze periode. Aldus luidt de derde algemene propositie: hoe hoger de (ervaren) concurrentie op de arbeidsmarkt in een bepaalde periode is, des te sterker steunt men seksdiscriminatie op de arbeidsmarkt. Er zijn minstens twee maatschappelijke condities welke de ervaren concurrentie op de arbeidsmarkt beïnvloeden. Wanneer het werkloosheidspercentage hoog is, zullen meer mensen moeten concurreren om hetzelfde aantal hulpbronnen op de arbeidsmarkt. Hierdoor zal de (ervaren) concurrentie toenemen. Hetzelfde geldt wanneer de mate van economische groei gering is, hetgeen leidt tot de volgende hypothesen met betrekking tot periode effecten: *(H3a) Hoe hoger het werkloosheidspercentage en (H3b) hoe geringer het percentage economische groei, des te meer steunt men seksdiscriminatie op de arbeidsmarkt in Nederland tussen 1979 en 1994.*

Niet alleen huidige omstandigheden kunnen van invloed zijn op de mate waarin concurrentie op de arbeidsmarkt ervaren wordt. Er kunnen eveneens invloeden uitgaan van de periode waarin men opgroeide, de zogenaamde formatieve jaren (cf. Mannheim, 1964). Wanneer we dit idee gebruiken in combinatie met dat wat uit de Conflict Theorie afgeleid kan worden, komen we tot de vierde algemene propositie: hoe hoger de ervaren concurrentie op de arbeidsmarkt was gedurende de formatieve levensfase, des te sterker steunt men seksdiscriminatie op de arbeidsmarkt. Wederom lijken de eerdergenoemde condities van toepassing te zijn. Wanneer tijdens de formatieve periode minder schaarse hulpbronnen verdeeld konden worden over hetzelfde aantal mensen, bijvoorbeeld in tijden van hoge werkloosheid of bij een geringe economische groei, mogen we verwachten dat door de mensen die aan dergelijke omstandigheden zijn blootgesteld, in het heden een hogere mate van concurrentie op de arbeidsmarkt ervaren wordt. Zodoende komen we tot de volgende toetsbare hypothesen over cohorteffecten: *(H4a) hoe hoger het werkloosheidspercentage tijdens de formatieve levensfase en (H4b) hoe geringer de economische groei was gedurende de formatieve levensfase, des te meer steunt men in het heden seksdiscriminatie op de arbeidsmarkt.*

Wanneer men leeftijds-, periode- en cohort effecten gezamenlijk opneemt in een model komt men voor een methodologisch probleem te staan. Dit leeftijd-periode-cohort identificatie probleem (Glenn, 1977; Mason et al. 1973) is in deze studie opgelost door zowel de periode-, als de cohort-effecten te vervangen door indicatoren die in theoretisch opzicht samenhangen met periode en cohort (Menard, 1991). Het probleem heeft zich nu verschoven van een methodologische naar een theoretische kwestie, omdat nu getoetst dient te worden of deze theoretisch relevante maatschappelijke kenmerken werkelijk dienst doen als substituten voor het periode- en/of cohort effect (Menard, 1991; de Graaf, 1999).

3. DATA EN MEETINSTRUMENTEN

De hypothesen worden getoetst aan de hand van de data ontleend aan het project ‘Culturele Veranderingen in Nederland’ van het Sociaal en Cultureel Planbureau. Vanaf 1979 tot en met 1994 zijn er, enkele tussenliggende jaren uitgezonderd, items aanwezig uit welke de afhankelijke variabele “steun voor seksdiscriminatie op de arbeidsmarkt” geconstrueerd kan worden. In de jaren waarin deze items aan de respondenten voorgelegd werden, is de manier van vragen steeds hetzelfde geweest. De steekproeven van de afzonderlijke jaren zijn getrokken door aan de hand van urbanisatiegraad gemeenten in Nederland te selecteren. Uit deze gemeenten zijn een aantal wijken aselekt getrokken, waaruit per interviewer een startadres getrokken is, van waaruit deze interviewer volgens een voorgeschreven patroon mensen binnen deze wijk contacteert. Omdat de steekproeftrekking in hoge mate aselekt geschiedt en omdat het aantal respondenten per jaar voldoende hoog is, beschouwen we de afzonderlijke jaren en de geaggregeerde dataset als representatief voor de Nederlandse bevolking van 16 tot en met 74 jaar. In ieder afzonderlijk jaar van meting telt de steekproef ongeveer 2000 respondenten.

Afhankelijke variabele

De afhankelijke variabele “steun voor seksdiscriminatie op de arbeidsmarkt” is geconstrueerd uit een tweetal items, waarin aan de respondent gevraagd wordt om in twee specifieke situaties een keuze te maken tussen een man of een vrouw. Deze twee situaties zijn omschreven als het vergeven van een baan, of het geven van een promotie aan een man, dan wel een vrouw. Zowel de man als de vrouw beschikken in deze situaties over dezelfde capaciteiten en kwalificaties. De respondenten konden drie geldige antwoorden geven, namelijk “aan de man”, “aan de vrouw” en “mag geen verschil maken”. Deze laatste antwoordcategorie is niet aan de respondenten voorgelegd, om sociale wenselijkheid tegen te gaan. Uit geaggregeerde gegevens (niet gepresenteerd) blijkt dat veruit het hoogste percentage respondenten aangeeft dat het geen verschil mag maken. De items vertonen een correlatie van 0,48 met elkaar. Deze correlatie is significant en als betrekkelijk hoog te classificeren. Wanneer we naar de Cronbach’s alpha kijken, blijkt deze met 0,78 voldoende hoog uit te vallen. De geconstrueerde nominale schaal bestaat uit een nul- en een één categorie. De respondenten die zich in categorie 1 bevinden hebben bij minstens één van de items aangegeven de voorkeur voor de man boven de vrouw op de arbeidsmarkt te geven.

Onafhankelijke variabelen

In het theoretische gedeelte kwam naar voren dat met betrekking tot een vijftal individuele kenmerken hypothesen getoetst zullen gaan worden. Deze zullen hieronder beschreven worden. Ook zullen zowel periode als cohort kenmerken als onafhankelijke variabelen beschreven worden. Het eerste individuele kenmerk *geloofsgemeenschap* bestaat uit vijf categorieën die in de tabellen te vinden zijn. Als controle variabele zal ook de mate van *kerkgang* opgenomen worden. Het individuele kenmerk *geboortecohort* is van een continu kenmerk teruggebracht naar een zevental categorieën, beginnend met “1905 tot en met 1914” en eindigend met “1965 tot en met 1978”. De tussenliggende categorieën beslaan allen een range van tien geboortejaren. De variabele *opleidingsniveau* bestaat uit zeven categorieën, lopend van basisonderwijs tot een wetenschappelijke opleiding. *Sociale positie* is ingedeeld in een tiental categorieën die

eveneens in de tabellen te vinden zijn. Tot slot zal de variabele *geslacht* opgenomen worden in de analyses.

Voor de periode-, en cohortkenmerken zal het *werkloosheidspercentage*, het *percentage economische groei* en het *percentage vrouwen op de arbeidsmarkt* gebruikt worden. Deze tijdreeksen zijn aan de hand van gegevens van het Centraal Bureau voor de Statistiek gecreëerd (CBS, 1994). Voor de periodekenmerken heeft iedere respondent het betreffende percentage toegewezen gekregen van het jaar voordat hij of zij ondervraagd is. Waar voor enkele meetpunten gegevens ontbraken, is door middel van intrapolatie alsnog de tussenliggende waarde berekend. Aldus hebben we gegevens voor ieder meetmoment in de periode van 1979 tot en met 1994. Met betrekking tot de werkloosheid een jaar voor de meting, lopen de percentages uiteen van 11,2 procent in 1985 tot 3,9 procent in 1980. Het percentage economische groei is gemeten door de volumemutatie ten opzichte van het vorig jaar op te nemen. Deze mutaties lopen uiteen van -1,4 procent in 1983, tot 4,1 procent in 1991. Het percentage vrouwen op de arbeidsmarkt is een lineair stijgende tijdreeks. In 1979 nam 27,0 procent van de vrouwen deel aan het arbeidsproces, in 1994 is dit percentage gestegen tot 37,3 procent.

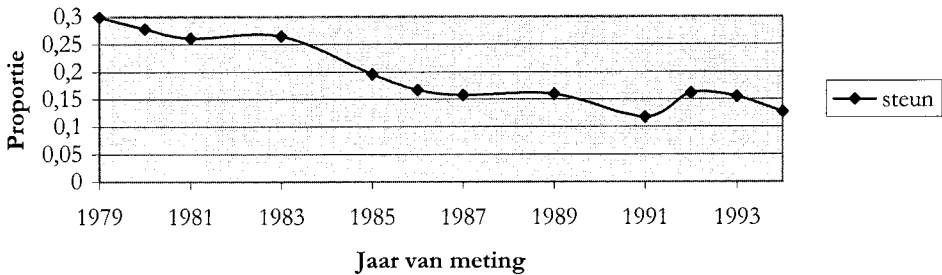
Voor de cohortkenmerken is dezelfde procedure gebruikt, met het verschil dat de percentages gekoppeld zijn aan het jaar waarin de respondent 18 jaar oud was. Zodoende beschikken we over tijdreeksen die betrekking hebben op de geboortejaren vanaf 1905 tot en met 1978. Alleen voor het werkloosheidspercentage geldt dat het berekend is door het gemiddelde percentage te berekenen in de periode dat de respondent tussen de 16 en de 20 jaar was (vgl. Coenders & Scheepers, 1998). Met betrekking tot het percentage vrouwen op de arbeidsmarkt geldt dat dit percentage tussen 1905 en 1955 slechts licht gestegen is. Daarna heeft de stijging van dit percentage zich in een hoger tempo doorgezet.

4. ANALYSE EN RESULTATEN

Om de verschillende hypothesen te toetsen, zullen twee verschillende regressietechnieken gebruikt worden. De individuele socialisatie hypothesen zullen getoetst worden door gebruik te maken van regressie met nominale variabelen oftewel RENOVA (Pelzer & Lammers, 1991). De maatschappelijke socialisatie hypothesen en de hypothesen met betrekking tot de periode-, en cohortkenmerken zullen met behulp van logistische regressie aan de empirie onderworpen worden. De RENOVA-techniek, ontwikkeld voor regressie analyse op nominale variabelen, wordt gebruikt omdat de individuele socialisatie hypothesen zodanig geformuleerd zijn, dat de referentiecategorie het algemeen gemiddelde van de steun voor seksediscriminatie op de arbeidsmarkt in een bepaald jaar is. Feitelijk is dat de proportie van de steekproef die seksediscriminatie steunt. Met de gebruikte analysemethode wordt automatisch dit algemeen gemiddelde als referentiecategorie behandeld. Een voordeel hiervan is dat de resultaten uitermate goed te vergelijken zijn met de trend in steun voor seksediscriminatie op de arbeidsmarkt. Immers, deze trend bestaat uit de algemene gemiddelden van de afzonderlijke jaren. Een ander voordeel is dat op deze wijze specifieke categorieën kunnen worden vergeleken met “de gemiddelde Nederlander”. Voor de toetsing van de overige hypothesen gebruiken we logistische regressie.

De toetsing van de individuele socialisatie hypothesen

In figuur 1 staat de algemene trend in de steun voor seksdiscriminatie op de arbeidsmarkt weergegeven. Uit deze figuur blijkt dat men tussen 1979 en 1994 steeds minder steun is gaan vertonen voor seksdiscriminatie op de arbeidsmarkt. In dit tijdsbestek is het percentage gedaald van ongeveer 30 procent naar ongeveer 13 procent. De trend is niet altijd monotoon dalend geweest, zowel tussen 1981 en 1983 als tussen 1991 en 1992 is er een lichte stijging in de steun voor seksdiscriminatie op de arbeidsmarkt waar te nemen.



Figuur 1. Steun voor seksdiscriminatie op de arbeidsmarkt, de proportie in het jaar van meting, leeftijd 18-74, N=23469.

Dan komen we nu toe aan de beschrijving van de resultaten met betrekking tot de hypothesen over de individuele socialisatie. Door RENOVA te gebruiken, komen de constanten van de regressievergelijkingen van de verschillende jaren overeen met de proporties per jaar die afgebeeld staan in figuur 1. In tabel 1 staan de resultaten van de toetsing van hypothese 1a tot en met 1j. In deze tabel staan per jaar de afwijkingen van het gemiddelde van de categorieën die behoren tot de variabelen “geloofsgemeenschap”, “geboortecohort”, “opleidingsniveau”, “sociale positie” en “geslacht”. Om de juiste afwijkingen te kunnen schatten voor geloofsgemeenschap is eveneens gecontroleerd voor de mate van kerkgang. Nemen we de parameter van $-0,17$ in 1979 van de mensen die een wetenschappelijke opleiding voltooid hebben (WO), dan betekent dit dat deze sociale categorie het algemeen gemiddelde van $0,31$ minus de betreffende parameter van $0,17$ is $0,14$ scoort op steun voor seksdiscriminatie op de arbeidsmarkt. Dit betekent dat deze sociale categorie seksdiscriminatie op de arbeidsmarkt significant minder steunt dan de “gemiddelde Nederlander” in het betreffende jaar, gecontroleerd voor alle kenmerken opgenomen in de vergelijking.

Hypothesen 1a tot en met 1c luidden dat respectievelijk gereformeerden, katholieken en Nederlands hervormden tussen 1979 en 1994 seksdiscriminatie op de arbeidsmarkt meer steunen dan de gemiddelde Nederlander. Gereformeerden wijkten slechts in twee jaren significant af van het algemeen gemiddelde. In 1986 was het verschil $0,09$ en in 1993 $-0,11$. In de meeste overige jaren was de afwijking wel positief, zoals verwacht, maar niet significant. Strikt genomen mogen we hypothese 1a dus niet aannemen, maar minder strikt genomen zijn de resultaten eerder in overeenstemming dan in tegenspraak met de hypothese. Voor de katholieken geldt hetzelfde. Hoewel in alle jaren, op één na, de verschillen volgens verwachting positief zijn, is de afwijking in slechts drie jaren significant positief, te weten 1980, 1986 en 1992. Strikt genomen mogen we hypothese

Tabel 1. Resultaten RENOVA, steun voor seksdiscriminatie op de arbeidsmarkt.

Variabele	1979	1980	1981	1983	1985	1986	1987	1989	1991	1992	1993	1994
Algemeen gemiddelde	0,31*	0,27*	0,26*	0,27*	0,20*	0,16*	0,16*	0,16*	0,12*	0,16*	0,16*	0,13*
Geloofsgemeenschap												
Gcen	-0,04	-0,05*	-0,03	-0,01	-0,01	-0,03*	-0,04*	0,01	-0,02*	-0,02*	0,01	-0,01
Rooms Katholiek	0,04	0,06*	0,00	0,03	0,01	0,04*	0,03	-0,01	0,03	0,04*	0,02	0,03
Nederlands Hervormd	0,01	0,03	0,04	-0,05*	-0,01	0,00	0,08*	0,03	0,02	-0,01	0,06*	0,01
Gereformceerd	0,03	0,04	0,07	0,05	0,01	0,09*	0,05	-0,06	0,02	0,06	-0,11*	-0,03
Ander geloof	0,06	0,07	0,13*	-0,01	0,01	0,01	-0,00	-0,08	0,02	0,07	-0,12*	0,04
Kerkgang												
Vaak	0,06*	0,01	0,03	0,03	0,12*	0,01	0,05*	0,09*	0,03	0,00	0,11*	0,05
Zelden	-0,02	-0,04	-0,02	0,02	-0,01	-0,01	-0,02	0,00	-0,01	-0,00	-0,01	0,01
Nooit	-0,02	0,01	-0,00	-0,02	-0,03*	0,00	-0,00	-0,03*	-0,00	0,00	-0,02*	-0,01
Geboortecohort												
1905 tot en met 1914	0,08*	0,15*	0,18*	-0,05	0,21*	0,01	0,09	(nvt)	(nvt)	(nvt)	(nvt)	(nvt)
1915 tot en met 1924	-0,00	0,03	0,14*	-0,01	0,03	0,01	0,07*	0,07	0,07*	0,09*	0,17*	0,10*
1925 tot en met 1934	-0,02	-0,05*	0,01	0,01	-0,02	0,06*	-0,02	0,02	0,07*	0,04	0,01	0,04
1935 tot en met 1944	-0,01	-0,02	-0,03	0,04	0,03	0,02	-0,00	-0,01	-0,01	0,01	0,03	0,00
1945 tot en met 1954	-0,04*	-0,03	-0,07*	-0,01	-0,05*	-0,05*	-0,02	-0,01	-0,02	-0,03*	-0,01	-0,01
1955 tot en met 1964	0,03	-0,01	-0,08*	-0,01	-0,00	-0,02	-0,02	-0,02	-0,05*	-0,02	-0,05*	-0,01
1965 tot en met 1978	(nvt)	(nvt)	0,02	0,02	-0,03	0,03	0,01	-0,01	0,02	-0,01	0,00	-0,03
Opleiding												
Basisschool	0,10*	0,15*	0,06*	0,19*	0,12*	0,13*	0,15*	0,16*	0,10*	0,12*	0,10*	0,09*
LBO	0,05*	0,08*	0,10*	0,06*	0,07*	0,04*	0,06*	0,04*	0,06*	0,07*	0,06*	0,07*
MULO, MAVO	-0,00	-0,03	-0,02	-0,05	-0,03	-0,04	-0,01	-0,01	-0,01	0,03	-0,01	0,00
MBO	-0,03	-0,05*	-0,03	0,02	-0,05*	-0,01	-0,03	-0,03	-0,04*	-0,05*	0,02	-0,02
HAVO, VWO	-0,14*	-0,13*	-0,11*	-0,12*	-0,04	-0,05*	-0,08*	-0,05	-0,05*	-0,05	-0,06*	-0,02
HBO	-0,15*	-0,12*	-0,10*	-0,14*	-0,10*	-0,08*	-0,04*	-0,09*	-0,04*	-0,07*	-0,07*	-0,07*
WO	-0,17*	-0,20*	-0,15*	-0,19*	-0,10*	-0,10*	-0,10*	-0,09*	-0,05*	-0,09*	-0,08*	-0,07*
Sociale positie												
Werkdoo	-0,19*	0,03	-0,04	0,01	-0,08*	0,01	-0,05	-0,04	0,13*	0,03	-0,01	-0,01
Huishouden	0,05*	0,02	0,04*	0,03	0,02	0,02	0,03	0,03*	0,00	-0,00	0,03	0,03
Uitkeringsgerechtigd	0,04	-0,01	0,00	0,03	0,00	-0,02	0,01	-0,04	0,03	0,01	0,02	0,03
Scholier, student	-0,06	-0,01	-0,00	-0,03	-0,02	-0,08	-0,05	0,02	-0,00	-0,02	-0,04	-0,00
Gepensionceerd	0,04	0,03	-0,05	0,07	0,02	0,00	0,03	-0,03	-0,05	0,05	0,02	0,01
Loondienst	-0,06*	-0,03	-0,01	-0,04	0,00	0,01	-0,01	-0,01	0,01	-0,01	-0,03*	-0,02
Overheidsdienst	-0,02	-0,03	-0,06*	-0,05	-0,00	-0,01	-0,04	-0,02	-0,02	0,01	-0,01	-0,02
Zelfstandig	0,06	0,02	0,03	0,03	-0,01	-0,03	0,08*	-0,04	-0,01	0,00	0,09*	0,03
Anders	0,01	0,01	-0,01	-0,07	-0,03	-0,05	-0,06	-0,06	-0,03	-0,01	-0,02	-0,03
Geslacht												
Man	0,04*	0,03*	0,04*	0,03*	0,01	0,02*	0,02*	0,02	0,01	0,01	0,02	0,02
Vrouw	-0,01	-0,03*	-0,04*	-0,03*	-0,01	-0,02*	-0,02*	-0,01	-0,01	-0,01	-0,01	-0,01
Aantal	1861	1870	1890	1803	1881	1867	1828	1769	1711	1875	1664	1748

Bron: Sociaal en Cultureel Planbureau, Culturele Veranderingen 1979-1994, * p < 0,05

Ib dus niet aannemen. In 9 van de 12 jaren is de afwijking van het algemeen gemiddelde voor Nederlands hervormden volgens verwachting positief. Slechts in 2 van deze 9 gevallen is de afwijking ook daadwerkelijk significant, namelijk in 1987 en 1993. In 1983 is de afwijking tegen de verwachting in significant negatief. Ook hier kunnen we hypothese 1c strikt genomen niet aannemen.

Deze kleine en soms inconsistente verschillen zijn wellicht te wijten aan de mate van kerkgang, die immers preciezer aangeeft in hoeverre men is blootgesteld aan respectievelijk zich verwant voelt met de cultureel conservatieve ideologie. Uit de resultaten blijkt dat degenen die vaak ter kerke gaan inderdaad seksdiscriminatie op de arbeidsmarkt meer steunen: alle afwijkingen ten opzichte van de gemiddelde Nederlander zijn positief en in 5 van de 12 jaren zijn deze afwijkingen significant.

Hypothese 1d luidde dat personen die geboren zijn voor 1940 seksdiscriminatie op de arbeidsmarkt meer steunen dan de gemiddelde Nederlander tussen 1979 en 1994. Wanneer we de resultaten in tabel 1 beschouwen, zien we dat deze sterk in het voordeel van deze hypothese spreken. Wanneer we de drie oudste geboortecohorten bekijken, zien

we slechts in 1980 een significant negatieve afwijking van het algemeen gemiddelde voor de personen die tussen 1925 en 1934 geboren zijn. De overige significante afwijkingen zijn allen positief en veelal sterk te noemen. Vooral voor het oudste geboortecohort, geboren tussen 1905 en 1914, zien we betrekkelijk sterke afwijkingen tussen 1979 en 1985. Voor het één na oudste geboortecohort, geboren tussen 1915 en 1924, zien we vanaf 1987 grote positieve afwijkingen. De jongere cohorten, geboren tussen 1945 en 1954 respectievelijk tussen 1955 en 1964, vertonen daarentegen minder steun voor seksdiscriminatie: op een enkele uitzondering na zijn de afwijkingen van de gemiddelde Nederlander negatief en in een aantal gevallen zijn de afwijkingen significant. Het jongste cohort, geboren tussen 1965 en 1978, vormt eventueel een uitzondering: zij wijken nauwelijks en in geen enkel jaartal significant af van de gemiddelde Nederlander. In vergelijking met de twee cohorten voor hen zou dat beschouwd kunnen worden als een trendbreuk. Hier komen we in de discussie nog op terug. Het blijkt inderdaad zo te zijn dat personen geboren voor 1940 seksdiscriminatie op de arbeidsmarkt meer steunen tussen 1979 en 1994.

Hypothese 1e en 1f luidden dat respectievelijk personen zonder een diploma en personen met een LBO-diploma tussen 1979 en 1994 seksdiscriminatie op de arbeidsmarkt meer steunen dan de gemiddelde Nederlander. Beide hypothesen mogen niet verworpen worden. Voor alle jaren zijn de afwijkingen van het algemeen gemiddelde voor beide opleidingscategorieën significant positief. Verder blijkt dat in alle jaren op één jaar na, namelijk in 1981, personen die alleen de basisschool hebben afgerond seksdiscriminatie op de arbeidsmarkt sterker steunen dan personen met een LBO-diploma.

Hypothesen 1g tot en met 1i luidden dat respectievelijk ongeschoolde handarbeiders, huisvrouwen en kleine zelfstandigen tussen 1979 en 1994 seksdiscriminatie op de arbeidsmarkt meer steunen dan de gemiddelde Nederlander. Hypothese 1g kunnen we niet toetsen, omdat de data het niet toelaten om een beroepenindeling te maken. Slechts de sociale positie van de respondent is bekend. De enige categorie die een beetje in de buurt komt zijn de personen in loondienst. Wanneer we de resultaten voor deze categorie bekijken moeten we tot de conclusie komen dat hypothese 1g niet aangenomen mag worden, wat niet verwonderlijk is, omdat deze categorie verre van homogeen is. Hypothese 1h mogen we niet verwerpen. Huisvrouwen blijken in alle jaren op één jaar na een positieve afwijking van het algemeen gemiddelde te hebben. In drie jaren is deze afwijking ook daadwerkelijk significant. Hoewel de afwijkingen veelal klein zijn blijken huisvrouwen seksdiscriminatie meer te steunen dan de gemiddelde Nederlander. De resultaten voor de (kleine) zelfstandigen zijn minder eenduidig. Maar hoewel er enkele niet significante negatieve afwijkingen te vinden zijn, zijn de enige twee significante afwijkingen sterk positief. Op basis van deze gegevens mogen we hypothese 1i met enige voorzichtigheid aannemen.

De laatste hypothese met betrekking tot de individuele socialisatie (H1j) luidde dat mannen tussen 1979 en 1994 seksdiscriminatie op de arbeidsmarkt meer steunen dan de gemiddelde Nederlander. Deze hypothese mogen we aannemen. Hoewel de afwijkingen van het algemeen gemiddelde in de loop van de tijd kleiner zijn geworden en na 1987 ook niet meer significant zijn, zijn deze afwijkingen altijd volgens verwachting positief. Mannen steunen seksdiscriminatie op de arbeidsmarkt aldus meer dan de gemiddelde Nederlander, maar dit verschil is na 1987 onbeduidend geworden.

De toetsing van de maatschappelijke socialisatie hypothesen

In tabel 2 staan de resultaten van de logistische regressie weergegeven. Van belang in

Tabel 2. Steun voor seksdiscriminatie op de arbeidsmarkt, leeftijd 18-24 jaar.

	Model 1		Model 2	
Periode				
Werkloosheid	-0,036**	(0,007)	-0,033**	(0,007)
Economische groei	-0,075**	(0,013)	-0,091**	(0,013)
Percentage vrouwen	-0,072**	(0,005)	-0,084**	(0,006)
Cohort				
Werkloosheid	0,065**	(0,007)	0,010	(0,009)
Economische groei	-0,008	(0,007)	0,001	(0,007)
Percentage vrouwen	-0,154**	(0,007)	0,023	(0,015)
Denominatie				
Rooms Katholiek			0,289**	(0,060)
Nederlands Hervormd			0,248**	(0,065)
Gereformeerd			0,302**	(0,084)
Ander geloof			0,255*	(0,110)
Geen geloof (ref. cat)				
Kerkgang				
1 keer per week of vaker (ref. cat)				
Eens in de twee weken			-0,360**	(0,088)
Maandelijks			-0,260**	(0,081)
Minder dan 1 keer per maand			-0,295**	(0,063)
Nooit			-0,353**	(0,065)
Geslacht (vrouw)			-0,334**	(0,052)
Opleiding			-0,297**	(0,012)
Sociale positie				
Werkloos			0,145	(0,123)
Huishouden			0,414**	(0,082)
Uitkeringsgerechtigd			0,116*	(0,103)
Student/scholier			0,138	(0,112)
Gepensioneerd			0,343**	(0,096)
In loondienst			0,116	(0,074)
Zelfstandig			0,385**	(0,103)
Andere sociale positie			-0,060	(0,167)
In overheidsdienst (ref. cat)				
Leeftijd				
			0,015**	(0,002)
Constante	4,992**		1,737**	
Aantal	23078		21742	

Bron: Sociaal Cultureel Planbureau, Culturele Veranderingen 1979-1994

* significant $p < 0,05$ ** significant $p < 0,01$

deze tabel zijn de periode-, en cohortkenmerken die betrekking hebben op de maatschappelijke socialisatie hypothesen en de hypothesen over de economische kenmerken (Model 1). Deze kenmerken zijn gecontroleerd voor de individuele kenmerken van de respondent (Model 2). In de tabel staan de b-coëfficiënten afgebeeld, met daarachter tussen haakjes de standaardfouten.

Allereerst beschouwen we de coëfficiënten die betrekking hebben op de hypothesen over de maatschappelijke socialisatie. Uit de significante b-waarde van $-0,072$ in model 1 blijkt dat men seksdiscriminatie op de arbeidsmarkt minder steunt wanneer er op het moment van meting meer vrouwen aan het arbeidsproces deelnemen. Deze invloed blijft van kracht wanneer in model 2 voor relevante individuele kenmerken gecontroleerd is. Volgens de Critical Mass Theory zou het sterkste effect kenbaar moeten worden wanneer de "critical mass" een percentage van 30 procent bereikt. Dit percentage werd bereikt tussen 1983 en 1985. Wanneer we figuur 1 nogmaals beschouwen, zien we dat de sterkste daling van de trend inderdaad tussen 1983 en 1985 te zien is. Deze bevinding kan opgevat worden als een extra sterke bevestiging van hypothese 2a. De significante b-waarde van $0,154$ in model wijst uit dat ook gedurende de forma-

tieve jaren het percentage vrouwen op de arbeidsmarkt de verwachte invloed vertoont. Echter, wanneer in model 2 gecontroleerd wordt voor individuele kenmerken, verdwijnt deze significante invloed. Hypothese 2b mag dus niet aangenomen worden. Wanneer een hoger percentage vrouwen deelnam aan het arbeidsproces in de tijd dat de respondent 18 jaar oud was, heeft dit niet de verwachte negatieve invloed op de steun voor seksdiscriminatie op de arbeidsmarkt.

In de paragraaf waar de invloed van macro-economische kenmerken besproken werd, is de verwachting uitgesproken dat mensen op traditionele waarden terugvallen wanneer zich een economisch minder voorspoedige tijd voordoet. Ook wanneer men opgegroeid is in een tijd van economische tegenslag ondersteunt men in het heden sterker traditionele waarden. Uit de invloed die het percentage werkloosheid heeft op de steun voor seksdiscriminatie op de arbeidsmarkt blijkt dat dit niet het geval is. Hoe hoger het percentage werkloosheid is een jaar voor het jaar van meting, zowel in model 1 als in model 2, des te minder steunt men seksdiscriminatie op de arbeidsmarkt. Hiermee moet hypothese 3a verworpen worden. Wanneer we naar het percentage economische groei een jaar voor het jaar van meting kijken, blijkt het effect in de verwachte richting te wijzen in zowel model 1 als 2. Hoe hoger het percentage economische groei is, des te minder steunt men seksdiscriminatie op de arbeidsmarkt. Hypothese 3b mag aldus niet verworpen worden. Met betrekking tot de macro-economische kenmerken in de formatieve jaren van de respondent zijn geen significante effecten aangetroffen. Dit betekent dat zowel hypothese 4a als 4b verworpen dienen te worden. Het percentage werkloosheid in de tijd dat de respondent tussen de 16 en 20 jaar oud was, noch het percentage economische groei in de formatieve jaren vertoont de respectievelijk verwachte positieve en negatieve invloed.

Voorts merken we op dat in model 2 de hypothesen 1a tot en met 1c inzake verschillen tussen denominaties niet verworpen worden: de leden van kerkgenootschappen steunen seksdiscriminatie meer dan niet-leden. Met betrekking tot kerkgang valt op dat de kerkleden die wekelijks naar de kerk gaan meer steun voor seksdiscriminatie vertonen dan enige andere categorie.

5. CONCLUSIE EN DISCUSSIE

De eerste centrale vraagstelling was een trendvraag en luidde: in hoeverre is er in Nederland tussen 1979 en 1994 sprake van een neerwaartse trend wanneer het gaat om de steun voor seksdiscriminatie op de arbeidsmarkt? Tussen 1979 en 1994 is de Nederlander inderdaad steeds minder seksdiscriminatie op de arbeidsmarkt gaan steunen. In 1979 steunde nog 31 procent van de Nederlanders seksdiscriminatie op de arbeidsmarkt, in 1994 was dit percentage inmiddels teruggelopen tot 13 procent.

De tweede centrale vraagstelling luidde: welke sociale categorieën steunen in Nederland tussen 1979 en 1994 seksdiscriminatie op de arbeidsmarkt?. Onze veronderstelling luidde dat steun voor seksdiscriminatie is ingegeven door de opvatting dat vrouwen een traditionele rol dienen te vervullen. Deze opvatting is ingebed in de cultureel conservatieve ideologie. Het bleek inderdaad zo te zijn dat de mensen die gesocialiseerd zijn in instituties die traditionele waarden doorgeven, of in tijden waarin traditionele waarden wijder verbreid waren, vaker seksdiscriminatie op de arbeidsmarkt steunden. Vooral mensen geboren voor de Tweede Wereldoorlog, laag opgeleiden, huisvrouwen en kleine zelfstandigen bleken steun te betonen voor seksdiscriminatie

op de arbeidsmarkt. Dit was ook, zij het in mindere mate, het geval voor mensen behorende tot een geloofsgemeenschap en voor mensen die veelvuldig ter kerke gaan.

Het sekseverschil in steun voor seksdiscriminatie op de arbeidsmarkt is in de loop der jaren verdwenen, maar een ander seksegerelateerd verschil is gebleven: huisvrouwen steunen nog steeds veel sterker dan de gemiddelde Nederlander seksdiscriminatie op de arbeidsmarkt. Meerdere verklaringen zouden hiervoor kunnen gelden. Het kan zo zijn dat de huisvrouwen menen dat de eigen positie de enig juiste is voor getrouwde vrouwen. Aangezien het aantal vrouwen dat de positie van huisvrouw inneemt is afgenomen (Hooghiemstra & Niphuis-Nell, 1993), kan bovendien een selectie-effect zijn opgetreden, waarbij vooral die vrouwen die een meer traditionele opvatting zijn toegegaan voor de positie van huisvrouw zijn blijven kiezen. Tenslotte zou er een verband kunnen zijn tussen de positie van huisvrouw en een relatieve isolatie van de maatschappij. Coser (1974) noemde het huisvrouwschap een “gulzige institutie”, daarmee aangevend dat de leden van die institutie (op basis van vrijwilligheid) in een positie van isolatie van de buitenwereld terechtkomen. Daarnaast willen we een opmerkelijke bevinding benadrukken: het cohort geboren na 1965 onderscheidt zich niet van de gemiddelde Nederlander in de steun voor seksdiscriminatie. Deze opmerkelijke bevinding krijgt meer reliëf wanneer men bedenkt dat dit jongste cohort meer geneigd is tot etnische discriminatie (vgl. Scheepers, 1996) alsook sterker geneigd is tot discriminatie van homoseksuelen (van de Meerendonk en Scheepers, 2002). Het heeft er alle schijn van dat dit jongste cohort een restauratie van traditionele normen verkiest (vgl. Scheepers et al., 1999).

Niet alleen op individueel vlak, maar ook op maatschappelijk vlak zijn de normen met betrekking tot deze traditionele rolverdeling veranderd. Daartoe is de derde centrale vraagstelling geformuleerd en deze luidde: welke periode-, en/of cohortkenmerken kunnen de neerwaartse trend in de steun voor seksdiscriminatie op de arbeidsmarkt verklaren? Onze propositie luidde dat in de loop van de tijd een geleidelijke toename van vrouwen op de arbeidsmarkt en daardoor gewenning te zien zou zijn wanneer het gaat over minder traditionele rolverdelingen. Deze toegenomen gewenning zou tot minder steun voor seksdiscriminatie op de arbeidsmarkt moeten leiden. Dit idee is voorts in een specifiekere theoretisch kader geplaatst door aan de hand van de “critical mass” theorie tot de stelling te komen dat een cultuurverandering plaatsvindt wanneer het percentage deelnemende vrouwen op de arbeidsmarkt in zo sterke mate toeneemt dat er een bepaalde zichtbaarheid en daardoor gewenning optreedt. Deze gewenning zou zowel voor de periode waarin men leeft, als voor de periode waarin men opgegroeid is kunnen gelden. Deze gedachten blijken ten dele te worden ondersteund door de resultaten. Het bleek inderdaad zo te zijn dat een toenemend percentage deelnemende vrouwen aan het arbeidsproces gedurende de periode geleid heeft tot een mindere steun voor seksdiscriminatie op de arbeidsmarkt. De neerwaartse trend in de steun voor seksdiscriminatie op de arbeidsmarkt loopt over de hele periode 1979-1994, maar is het sterkst in de jaren waarin het aantal vrouwen op de arbeidsmarkt de “critical mass” overschreden heeft, tussen 1983 en 1985. Moss Kanter’s theorie dat getalsverhoudingen in organisaties gevolgen hebben voor het gedrag van leden van die organisaties, lijkt daarmee ook op te gaan op het niveau van de samenleving als geheel. Er zijn ons geen andere studies bekend waarin haar theorie op het niveau van de samenleving met longitudinale data is getoetst. Het interessante van dit onderzoek is bovendien dat wij door het gebruik van longitudinale data konden laten zien dat het percentage van 30%, dat in navolging van Dahlerup als omslagpunt beschouwd wordt, ook werkelijk deze

functie lijkt te vervullen. Deze verwachtingen worden echter niet ondersteund aan de hand van de invloed die van de formatieve jaren uitgaat. Blijkbaar is de verandering in de norm over de traditionele rolverdeling tussen mannen en vrouwen iets van de laatste tijd en onafhankelijk van de tijd waarin men opgroeide. Deze bevinding vormt in zekere zin een uitzondering: ten aanzien van enkele andere aspecten van politieke opvattingen bleken er wel degelijk cohort-effecten te bestaan ten aanzien waarvan kon worden aangetoond dat die gerelateerd waren aan de omstandigheden tijdens de formatieve jaren (vgl. Coenders & Scheepers, 1998). Het feit dat dergelijke cohort-effecten ten aanzien van steun voor seksdiscriminatie niet bestaan, heeft wellicht te maken met het feit dat er juist ten aanzien hiervan een "brede maatschappelijke discussie" heeft gewoed, waardoor effecten gerelateerd aan de formatieve jaren worden "geneutraliseerd" door effecten van contemporaine aard.

Er kunnen echter maatschappelijke omstandigheden zijn die iemand traditionele waarden doet verkiezen. Dit zou het geval kunnen zijn in tijden van economische achteruitgang en voor mensen die opgegroeid zijn in economisch minder voorspoedige tijden. Deze verwachtingen zijn ingebed in de conflicttheorie die leidde tot de stelling dat de steun voor seksdiscriminatie op de arbeidsmarkt in een bepaalde periode beïnvloed wordt door de macro-economische omstandigheden in die periode en in de periode waarin men opgroeide. Voor de periode-effecten van het percentage economische groei bleek deze verwachting op te gaan, maar voor het percentage werkloosheid niet. Met betrekking tot de cohorteffecten zijn in het geheel geen significante invloeden gevonden. Het feit dat in de tijd waarin men leeft in ieder geval het percentage economische groei ervoor zorgt dat men seksdiscriminatie op de arbeidsmarkt minder steunt, versterkt de conclusie van de vorige alinea. Het zijn de huidige veranderingen in de maatschappelijke realiteit die ervoor zorgen dat men de ideeën over traditionele rolverdelingen aanpast, en niet zozeer de normen die men in de jaren dat men opgroeide meekreeg.

Samenvattend zien we tussen 1979 en 1994 een sterke daling in het percentage mensen dat seksdiscriminatie op de arbeidsmarkt ondersteunt. Het zijn in alle jaren vooral de personen die affiniteit hebben met de cultureel conservatieve ideologie, die seksdiscriminatie op de arbeidsmarkt steunen. In de loop van de tijd lijkt het normatieve tij met betrekking tot traditionele rolverdelingen tussen mannen en vrouwen gekeerd te zijn. Van groot belang daarbij lijkt te zijn dat de maatschappelijke realiteit een omslagpunt is gepasseerd tussen 1983 en 1985. Steeds meer mensen zijn gewend geraakt aan het feit dat steeds meer vrouwen aan het arbeidsproces deel zijn gaan nemen, hetgeen zich uitwijst in de afname van de steun voor seksdiscriminatie op de arbeidsmarkt.

LITERATUUR

- Berkel, M. van (1997). *Who dominates when. Asymmetrical patterns of influence among husbands and wives*. Nijmegen: Mediagroep.
- Blalock, H.M. (1967). *Toward a Theory of Minority Group Relations*. New York: Wiley.
- CBS (1994). *Vijfennegentig jaren statistiek in tijdreeksen, 1899-1994*. 'S Gravenhage: SDU Uitgeverij.
- Coenders, M. & Scheepers, P. (1998). Support for Ethnic Discrimination in the Netherlands 1979-1993: Effects of Period, Cohort, and Individual Characteristics. *European Sociological Review*, 14, 4, 405-422.
- Coser, L. (1956). *The Functions of Social Conflict*. Glencoe Illinois: Free Press.

- Coser, L. (1974). *Greedy institutions. Patterns of undivided commitment*. New York: Free Press.
- Dahlerup, D. (1988). From a Small to a Large Minority: Women in the Scandinavian Politics. *Scandinavian Political Studies*, 11,4, 274-298.
- Glenn, N.D. (1977). *Cohort Analysis*. Beverly Hills: Sage Publications.
- Hooghiemstra, B. T. J. & Niphuis-Nell, M. N. (1993). *Sociale atlas van de vrouw, deel 2. Arbeid, inkomen en faciliteiten om werk en de zorg voor kinderen te combineren*. Rijswijk/Den Haag: Sociaal en Cultureel Planbureau.
- Keuzenkamp, S. & Oudhof, K. (2000). *Emancipatiemonitor 2000*. Den Haag: Sociaal en Cultureel Planbureau.
- Lammers, J. & Pelzer, B. (1991). Regressie analyse met nominale variabelen. Achtergrond en beschrijving van een programma. Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.
- Lippe, T. van der & Doorne-Huiskes, J. van (1995). Veranderingen in stratificatie tussen mannen en vrouwen. Pp. 104-124 in Dronkers, J & Ultee, W.C. (red.). *Verschuivende ongelijkheid in Nederland. Sociale gelaagdheid en mobiliteit*. Assen: Van Gorkum en Comp. B.V.
- Mannheim, K. (1964). Das problem der generationen (The problem of generations). In Mannheim, K. (red.) *Wissensoziologie*. Berlin: Hermann Luchterhand Verlag
- Mason, K.O., Mason, W.M., Winsborough, H.H. & Poole, W.K. (1973). Some methodological issues in cohort analysis of archival data. *American Sociological Review*, 38, 242-258.
- Meerendonk, B. van de & Scheepers, P. (2001). Demand of equal rights for gay men and lesbians in the Netherlands 1980-1993. Verschijnt in *Journal of Homosexuality*.
- Menard, S. (1991). *Longitudinal Research*. Newbury Park: Sage Publications.
- Middendorp, C.P. (1991). *Ideology in Dutch politics, the democratic system reconsidered*. Assen, van Gorcum.
- Moss Kanter, R. (1977). *Men and Women of the Corporation*. New York: Basic Books Inc.
- Scheepers, P., Peters, J. & Felling, A. (2000). Individualisering van politieke opvattingen. Pp. 133-156 in Felling, A., Peters, J. & Scheepers, P. (red.) *Individualisering in Nederland aan het eind van de twintigste eeuw. Empirisch onderzoek naar omstreden hypotheses*. Assen: van Gorcum.
- Scheepers, P., Te Grotenhuis, M. & Bosch A. (1999). Trends in conservatism in the Netherlands 1970-1992. Pp. 25-50 in De Witte, H. & Scheepers, P. (red.). *Ideology in the low countries*. Assen: van Gorcum.
- Scheepers, P. (1996). Etnocentrische reacties en rechts-extremisme; reacties op de komst en aanwezigheid van etnische minderheden. Pp. 247-302 in Ganzeboom, H.G.B. & Ultee, W.C. (red.). *De sociale segmentatie van Nederland in 2015*. Den Haag: Sdu Uitgevers.
- Sociaal en Cultureel Planbureau (1994). *Sociaal en Cultureel Rapport 1994*. Rijswijk: Sociaal en Cultureel Planbureau.

11. De rol van marktonderzoek in crisis-communicatie

D. DE RAAF

SAMENVATTING

Voor marktonderzoek is in tijden van crisis geen tijd, zou men denken. De tijdsdruk is groot en veel doelgroepen vragen om antwoorden. De algemene PR boodschap is om goed voorbereid te zijn door een crisisplan te hebben en tijdens de crisis open en eerlijk te communiceren. Als er door ondernemingen marktonderzoek wordt gedaan is dit om een vinger aan de pols te houden wat het imago betreft. Uit deze studie komt echter een essentiële rol voor marktonderzoek naar voren. Door onderzoek kan opgespoord worden wat het publiek en klanten zien als oorzaak voor de crises. De responsestrategie dient vervolgens gebaseerd te worden op deze gepercipieerde oorzaken.

1. INLEIDING: BANDENSPANNING

Op 6 augustus 2000 maakt bandenfabrikant Firestone bekend dat een aantal typen banden gratis verwisseld mogen worden bij erkende dealers in de Verenigde Staten en Canada. Het lijkt een vrij normale product recall, waar nauwelijks aandacht aan wordt besteedt. Totdat een van de grootste winkelketens in de Verenigde Staten, Sears Roebuck, verklaart een aantal soorten Firestonebanden van het schap te halen. Het verhaal gaat dat de rubberlagen van de banden zouden loslaten en de banden als gevolg daarvan klappen. Dodelijke ongelukken als gevolg. Vier dagen later roept het moederbedrijf Bridgestone wereldwijd Firestone banden van het type ATX, ATXII en Wilderness AT terug van de markt. Deze gebeurtenissen roepen een groot aantal vragen op; wat is er precies aan de hand, wat is de oorzaak, kan de schade beperkt worden, is er sprake van schuld of van gewoon een ongeluk.

In de media worden ongelukken met de banden breed uitgemeten. Het officiële standpunt is dat het probleem nog niet in kaart is gebracht en dat er geen defecten in het ontwerp en de productie is gevonden. Er wordt gesuggereerd dat achterstallig onderhoud, verkeerd gebruik en beschadigingen veel van de ongelukken zouden verklaren. Kortom in eerste instantie worden echte problemen ontkend. Tevens verklaart de fabrikant dat er met een te lage bandenspanning, wordt gereden. In een later stadium wordt als oorzaak aangedragen dat het sabotage zou zijn van ontevreden werknemers in een periode van langdurige stakingen. Wel moet zij toegeven dat de banden inderdaad bij de genoemde ongelukken zijn betrokken en dat 80% van de ongelukken in de Zuidelijke staten plaatsvindt. De banden worden voornamelijk gebruikt voor terreinwagens. Al

deze banden worden geproduceerd in de fabriek in Decatur, Illinois. De productie wordt stil gelegd en circa 6,5 miljoen banden worden teruggehaald. Een geschatte kostenpost van 850 miljoen gulden.

Al snel blijkt dat bij bijna alle ongelukken gaat om een Ford Explorer. De verklaringen van de pers, lopen uiteen van een productiefout van de banden, een te zware auto voor genoemde banden en een door Ford te lage geadviseerde bandenspanning. Amerikaanse veiligheidsinstituten brengen de Firestone band en de Ford Explorer in verband met 270 gevallen met klapband, waarvan 62 met dodelijke afloop. Steeds slaat hierbij de Ford Explorer op de kop. Deze onderzoeken betrekken Ford duidelijk in de crisis en maken duidelijk dat het hier toch daadwerkelijk om problemen met de producten van deze fabrikanten gaat. Daarbij gaan de CEO's over en weer via de pers met elkaar ruziën bij wie de schuld ligt. John Lampe, de een na hoogste functionaris van Firestone, beschuldigt Ford ervan een te lage bandenspanning te adviseren. Jacques Nasser van Ford beschuldigt Firestone al lang van de problemen met de banden te weten en Ford niet op de hoogte te hebben gesteld. John Lampe zegt dat het aan de Explorer ligt en niet aan de banden. En daar waar beide topmannen over en weer ruziën ontstaat een vruchtbare bodem voor geruchten en paniek. Terughaalacties uit het verleden worden er bij gehaald, rechtszaken worden aangespannen en concurrenten gooien olie op het vuur. Wat leek te starten als een vrij normale product recall groeit uit tot een enorme crisis. Het aandeel van Bridgestone halveert in waarde en de terughaalactie brengt een kostenpost mee van 1 miljard gulden. Ford legt de productie in een aantal fabrieken neer om de productievoorraad vrij te maken voor de omruil actie. Ook betaalt zij mee aan de schade van Firestone. Daardoor daalt de kwartaalwinst van Ford met 16%, van 1,11 miljard dollar in het vorige jaar naar 888 miljoen dollar.

Bovenstaande casus is een voorbeeld waarbij een verkeerde repons van een merkenfabrikant rondom een productincident een enorme crisis tot gevolg kan hebben. Vooral de onduidelijkheid, het speculeren en het ruziën over de oorzaak lijkt een belangrijke rol in de escalatie van het incident te hebben gespeeld. Dit artikel beschrijft waarom de vraag naar de oorzaak, verantwoordelijkheid en schuld, cruciaal is in de escalatie van een incident en een belangrijke bepalende factor is in de manier waarop een organisatie zou moeten reageren. Daaruit zal blijken dat marktonderzoek in tijden van een crisis heel belangrijk is.

2. EEN PRODUCTCRISIS

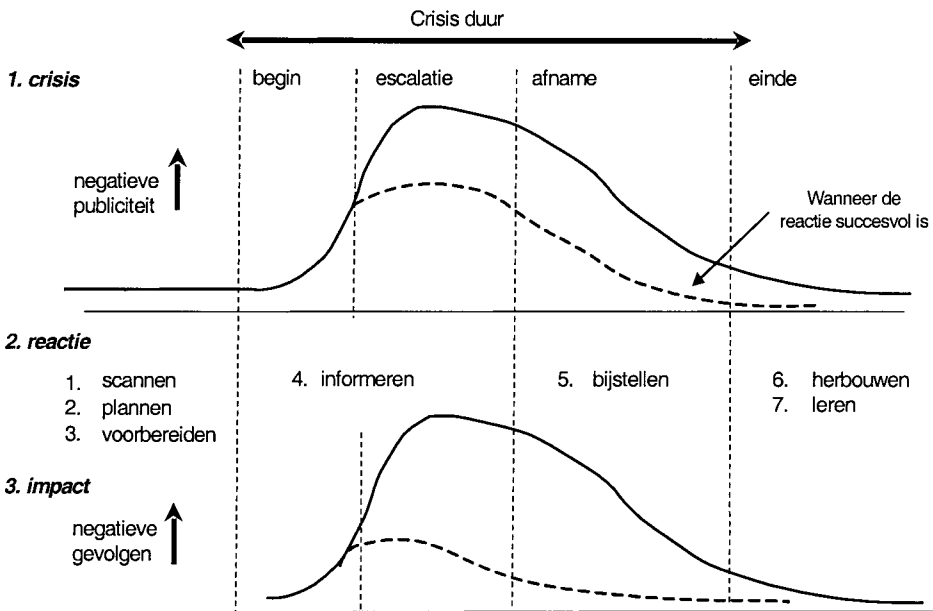
Een crisis kenmerkt zich in eerste instantie door schade voor de fabrikant, maar ook door een snelle escalatie van negatieve publiciteit, stress voor functionarissen en een zeer korte beslissingstijd. Er is een toename te bespeuren van dergelijke negatieve publiciteit in media over merken. De schade van een dergelijke crisis is vaak aanzienlijk. Behalve de kosten voor een eventuele recall en andere kosten verbonden aan de reactie, is er ernstige financiële schade verbonden aan de beschadiging van het imago van het merk. Het is moeilijk om aan imagobeschadiging een precies bedrag te koppelen, maar er wordt algemeen aangenomen dat de schade voor een bekend merk enorm is en de schattingen lopen in de miljoenen (Weinberger & Romeo, 1989; Kabak & Soimkos, 1992; Davidson & Worrell, 1992).

Wat doen bedrijven om in het geval van een crisis de schade zoveel mogelijk te beperken? Wees goed voorbereid, is een les die PR adviseurs ons in de afgelopen jaren goed hebben bijgebracht (Ten Berge, 1998; Birch, 1994; Mahoney, 1993). Deze crisisvoorbereiding houdt in de meeste gevallen in dat er een gedegen crisisplan in de kast ligt. Echter uit onderzoek (de Raaf, 1995) onder grote bedrijven in Nederland bleek dat slechts 47% een crisisplan daadwerkelijk voorhanden had. In deze plannen staat meestal beschreven: hoe de communicatielijnen lopen; wie er woordvoerder is; welke mensen ingeschakeld moeten worden; het proces om snel een call-centre op te richten; het proces om snel een persconferentie te beleggen en dergelijke. Het gaat eigenlijk voornamelijk om logistieke afspraken rondom het reageren op een crisis wanneer men met één geconfronteerd wordt.

3. CRISISCOMMUNICATIE

Minder goed voorbereid wordt wat de inhoud van de boodschap moet zijn. We kennen allemaal het rijtje dat we open, eerlijk en snel de feiten moeten communiceren. Maar in tijden van crisis biedt dit maar weinig houvast. In Figuur 1 staat weergegeven welk verloop een crisis in het algemeen heeft. De opbouw van een crisis (1. crisis) kent een aantal fasen die ieder een eigen reactie, een eigen vorm van communicatie behoeven (2. reactie). De schade neemt toe en af samen met de hoeveelheid negatieve publiciteit (3. impact).

In het begin, bij het uitbreken en de escalatie van een crisis, is het voornamelijk belangrijk om de consumenten te informeren (a) over hoe ze moeten reageren en (b) wat er nu



Figuur 1. Fasen in een merkgerelateerde crisis.

feitelijk aan de hand is. De nadruk ligt hier op feitelijk. Communiceer wat je echt weet en wat je echt doet als bedrijf. Speculeren is uit den boze. Bij het genoemde voorbeeld van Firestone hebben beide organisaties zich hier wel schuldig aangemaakt. Ondanks dat onderzoek geen uitsluitsel gaf over wat er nu precies de oorzaak van was, werden vele mogelijke oorzaken gesuggereerd.

Pas in een latere fase, gedurende de afbouw van de crisis dient de organisatie te proberen om de perceptie van de stakeholders bewust te beïnvloeden. Ook hier ziet men dat Firestone en Ford al meteen proberen de schade te beperken door de oorzaak buiten de eigen organisatie te leggen. In een later stadium zijn deze verklaringen niet meer houdbaar. Dit is zeer schadelijk voor het vertrouwen dat men in de organisaties heeft. Het vervolg van het artikel beperkt zich tot deze vorm van communicatie. Met welke boodschap kan je als organisatie in tijden van een crisis de beeldvorming over je organisatie zo positief mogelijk beïnvloeden.

In het kader van mijn promotieonderzoek zijn een drietal experimenten uitgevoerd waaruit blijkt dat het antwoord op de vraag waar de oorzaak van de crisis ligt, ofwel, wie er schuldig is, één van de belangrijkste verklaringen is voor de ernst van de gevolgen van een crisis. Ook het succes van de ondernemingsreactie was gelegen in de mate waarin de ondernemingsreactie aansloot op de oorzaak van het incident. De lange termijn gevolgen worden voornamelijk bepaald door de vraag of men het gebeurde toewijst aan de organisatie (veel merkbeschadiging) of meer aan de omgeving van de organisatie (minder merkbeschadiging).

4. DE ATTRIBUTIETHEORIE

Een bekend theoretisch kader voor het toewijzen van oorzaak-gevolg relaties in gebeurtenissen, is de attributietheorie (Heider, 1959; Shaver, 1985; Jaspars & Fincham, 1983; Kelley, 1976; Kelley & Michela, 1980). In de attributietheorie worden drie criteria onderscheiden waarmee consumenten proberen de oorzaak van een incident te begrijpen en oorzaken aan actoren of omstandigheden toe te schrijven. Aangepast aan de specifieke situatie van een merkgerelateerd incident zijn de volgende drie criteria:

(1) *Attributie van een interne of externe oorzaak (locus)* aan een incident reflecteert de vraag of de consument de oorzaak van het incident buiten de onderneming (externe attributie) of bij de organisatie zelf (interne attributie) legt.

Een voorbeeld waarbij de attributie in sterke mate extern was, is de vergiftiging van producten door een terroristische organisatie, zoals bij het merk Tylenol van Johnson & Johnson. De pijnstiller was in 1982 in de winkel ingespoten met cyanide. Hoewel de producent hier weinig invloed op had, leidde het Tylenol-incident tot overheidsregulering met betrekking tot sabotagebestendige productverpakkingen. Ondernemingen hebben bijna altijd wel enige beheersing over dergelijk incidenten, in de zin van voorzien en voorkomen. Maar in dit geval werd deze als miniem gepercipieerd, zoals bleek bij een onderzoek in 1986 onder consumenten, toen nogmaals een slachtoffer viel naar aanleiding van cyanide in een potje Tylenol-tabletten. Johnson & Johnson bleek nauwelijks last te hebben van omzetverlies naar aanleiding van de herhaling van vergifti-

ging van haar product. De consument oordeelde dat dergelijke terroristische acties van willekeur nooit geheel door een onderneming zijn uit te sluiten.

(2) Het tweede criterium in het attributieproces is de *beheersbaarheid* (controllability) van het gebeurde voor de onderneming. Het feit dat de oorzaak buiten de onderneming blijkt te liggen, neemt niet weg dat journalisten en consumenten dit toch als de verantwoordelijkheid van de onderneming zien, omdat de onderneming het had kunnen voorzien of voorkómen. Om een hoge mate van beheersbaarheid te percipiëren is het niet noodzakelijk dat er sprake is van intentie of opzet.

Bij de Olvarit crisis in 1993, waarbij de door Nutricia geproduceerde kindervoeding besmet leek te zijn met een chemische stof, Halamid genaamd, lag de herkomst van de besmetting van de babyvoeding bij een toeleverancier. Dit pleitte de onderneming echter niet vrij, omdat een onderneming immers zijn eigen leveranciers kiest en verantwoordelijk is voor de controle van de toegeleverde producten. Dus om een hoge mate van beheersbaarheid te percipiëren is het niet noodzakelijk dat er sprake is van intentie of opzet.

(3) “*Historie*” ook wel het *stabiliteitscriterium* genoemd, heeft betrekking op gelijksoortige gevallen in het verleden van de onderneming. Als een organisatie in het verleden verscheidene keren in de negatieve publiciteit is geweest, dan is de kans groter dat journalisten en consumenten de organisatie als schuldige aanwijzen, zelfs als volgens de beide andere criteria de schuld van de organisatie wordt verminderd.

De pers zal ook altijd op zoek gaan naar crises in het verleden van de organisatie. Bijkomend bij de Firestone crisis was dat een crisis uit het verleden waarin ook Firestone en Ford gezamenlijk in een crisis belandden, breeduit in de pers wederom werd uitgemeten. In 1978 was de Ford Pinto bij een pijnlijke crisis betrokken. De gemonteerde banden van Firestone bleken vaak last te hebben van klapbanden. Echter bij de vele aanrijdingen achterop, als gevolg van de klapband, ontplofte de benzine tank van de Ford Pinto. Deze was vlak achter de bumper gemonteerd. Het ging in dit geval om 41 dodelijke ongelukken. Ford was uiteindelijk genoodzaakt de productie te stoppen. Zeer pijnlijk aan deze crisis was dat er documenten naar buiten kwamen waaruit bleek dat Ford berekeningen had gemaakt op basis waarvan besloten was niet tot een product recall over te gaan. In deze berekeningen was geschat om hoeveel dodelijke ongelukken het zou gaan. Aan ieder leven werd een bedrag in de vorm van een schadevergoeding gekoppeld. Het uiteindelijke bedrag dat Ford kwijt zou zijn aan schadevergoedingen als gevolg van dodelijke ongelukken, was lager dan de berekende kosten van een product recall van alle Pinto modellen. Op basis hiervan is besloten om de auto niet terug te halen. Ook wordt bekend dat Ford in het Midden-Oosten al een jaar eerder een recall heeft gedaan, in verband met problemen met banden. Er waren voordelige banden onder de auto's gemonteerd, die niet bestand waren tegen de combinatie van de zinderende hitte in die omgeving en de hoge snelheden op lange brede snelwegen. Een paar maanden later werd Ford door een Californische rechtbank ook nog eens gedwongen om twee miljoen auto's terug te roepen voor reparatie vanwege problemen met de ontsteking, die te dicht bij motor zou zijn gemonteerd waardoor deze kan uitvallen.

De beschreven attributie van de oorzaak wordt niet alleen gezocht uit een soort nieuws-

gierigheid, maar vooral om de toekomst te kunnen voorspellen en adequaat te kunnen reageren.

In de tweede fase van een crisis is het de zaak om de attributie van de crisis zo voordelig mogelijk voor de onderneming te laten plaatsvinden. Gebaseerd op de attributietheorie, zijn een viertal typen crises te onderscheiden. Wanneer interne / externe attributie wordt gecombineerd met een hoge of een lage beheersbaarheid, krijgen we de volgende indeling van crises (zie Figuur 2). Een combinatie van de derde dimensie, stabiliteit levert geen andersoortige typen crises op en is daarom buiten deze typologie gelaten.

	Hoge beheersbaarheid	Lage beheersbaarheid
Externe attributie	<p style="text-align: center;">I</p> <p style="text-align: center;"><i>Incidenten in de directe omgeving van de organisatie</i></p> <p style="text-align: center;">verantwoordelijkheid / schuld schade -/+</p> <p>Bijvoorbeeld: Olvarit crisis, waarbij de oorzaak van de besmetting bij een toeleverancier lag</p>	<p style="text-align: center;">II</p> <p style="text-align: center;"><i>Terrorisme, geknoei met producten, rampen</i></p> <p style="text-align: center;">associatie / verantwoordelijkheid schade -/-</p> <p>Bijvoorbeeld: Tylenol crisis, waarbij onbekenden het product besmet hebben in de supermarkt</p>
Interne attributie	<p style="text-align: center;">III</p> <p style="text-align: center;"><i>overtredingen gevolgen te voorzien afkeurenswaardig gedrag</i></p> <p style="text-align: center;">schuld schade +/+</p> <p>Bijvoorbeeld: Ford die bewust goedkope banden monteert in het Midden-Oosten ondanks de bekende gevolgen van grote hitte</p>	<p style="text-align: center;">IV</p> <p style="text-align: center;"><i>ongelukken niet voorzien geen intentie</i></p> <p style="text-align: center;">verantwoordelijkheid / schuld schade +/-</p> <p>Bijvoorbeeld: Brinta waarbij een salmonella besmetting van een lopende band de reden was voor het terughalen van het ontbijtproduct.</p>

Figuur 2. Vier typen crises.

In het eerste experiment werd gekeken in welke mate manipulaties in attributiekennmerken van een crisis, ook dezelfde percepties bij de respondenten oproep en of deze consequenties had voor de gevolgen van een crisis. Er werd gekeken of manipulaties van crisisbeschrijvingen met betrekking tot de drie criteria van de attributietheorie tot verschillen in perceptie van de drie criteria leidden. Op basis van de volgende kenmerken, interne versus externe locus, hoge versus een lage beheersbaarheid en eerder in opspraak geweest ja versus nee, werden een zestal denkbeeldige crises beschreven. De respondenten in het experiment kregen een tweetal krantenartikelen voorgelegd waarin de crises werd beschreven. Vervolgens werd hen gevraagd een vragenlijst in te vullen. Er werd gevraagd: of zij de oorzaak binnen of buiten de organisatie vonden liggen (*locus*, 2 items, cronbach $\alpha = .96$), of ze vonden dat de organisatie beheersing had

gehad over het gebeurde (*beheersbaarheid*, 5 items, cronbach $\alpha = .70$), en of ze het gevoel hadden dat de kans dat dit bij deze organisatie gebeurde groter was dan gemiddeld (*stabiliteit*, 1 item). De respondentengroep bestond uit 44 studenten. Er waren een zestal experimentele groepen. De groepen bestonden dus uit slechts 6 tot 7 respondenten. De tweezijdige t-testen vonden plaats tussen twee groepen van ieder 22 respondenten. Toch waren de resultaten overtuigend. In onderstaande tabel staan de manipulaties met bijbehorende tweezijdige t-test met betrekking tot de schalen en variabelen die bedoeld waren om de manipulaties te meten.

Tabel 1. Manipulatie-check criteria attributie theorie, tweezijdige t-test tussen groepen in het experiment.

Perceptie (7-punts schalen) → Manipulatie ↓	Locus	Beheersbaarheid	Stabiliteit
Interne versus externe oorzaak	6,01 vs. 3,47 (.000)	4,77 vs. 3,65 (.001)	4,41 vs. 4,13 (.659)
Lage versus een hoge beheersbaarheid	6,00 vs. 6,11 (.772)	4,18 vs. 5,35 (.000)	4,46 vs. 4,36 (.887)
Eerder in opspraak geweest; nee versus ja	4,55 vs. 5,47 (.087)	4,28 vs. 4,36 (.835)	3,30 vs. 5,17 (.000)

De vetgedrukte cijfers in tabel 1 geven het verschil aan in gemiddelde score tussen twee manipulaties in het experiment. De score van locus (een hoge score betekent dat de mensen de oorzaak binnen de organisatie zoeken een lage score betekent dat men de oorzaak buiten de organisatie percipieert) verschilt significant tussen de manipulatie waarin in een krantenartikel was geschreven waarin de oorzaak meer in de organisatie zelf lag en een krantenartikel waarin de oorzaak meer buiten de organisatie lag. Dit theoretische onderscheid werd, zoals blijkt uit de significante verschillen, door de respondenten op dezelfde manier gezien. Alle verwachte effecten, de manipulatie met de beheersbaarheid en of de organisatie eerder in opspraak was geweest, hadden de verwachte effecten op de schalen die de gerelateerde perceptie diende te meten. In alle gevallen waren de effecten significant en in de juiste richting. Dit bevestigt dat de theoretische manipulaties vanuit de attributietheorie, overeenkomen met daadwerkelijke criteria die mensen gebruiken in het verwerken van informatie over een crisis.

5. CRISIS-RESPONSE STRATEGIEËN

In een ander experiment van het promotieonderzoek werd de typologie van figuur 2 gebruikt om na te gaan welke responsestrategie het beste was voor een bepaald type crisis. Er worden vijf crisisresponse strategieën onderscheiden (Coombs, 1995; Rossiter & Percy, 1995; Tybout & Calder, 1981; Van Raaij, 1987; Van den Bosch & Van Riel, 1996; Sturges, 1994; Ice, 1991; Wolfe, 1990).

Bij de *verwijsstrategie* ontkent de organisatie enige oorzaak relatie met het incident te hebben. Zij zegt eigenlijk: "ik heb er niks mee te maken" of "ik heb het niet gedaan". De *distancierstrategie* probeert de gepercipieerde beheersbaarheid te verminderen en omvat tactieken als: "ik wist het niet" of "ik kon er niets aan doen". De *herinnerings-*

Tabel 2. Crisis responsestrategieën.

-
1. Verwijsstrategie (externalisation strategy)
 2. Distantiestrategie (distance strategy)
 3. Herinneringsstrategie (retrieval strategy)
 4. Boetekleedstrategie (mortification strategy)
 5. Lijdstrategie (suffering strategy)
-

strategie brengt in herinnering het goede dat de organisatie heeft gedaan dan wel waar het merk voor staat. Het probeert op deze manier de negatieve associaties te compenseren door positieve associaties te benadrukken. In het geval van de *boetekleedstrategie* neemt de organisatie volledig de schuld en verantwoordelijkheid op zich. Bij de *lijdstrategie* probeert de organisatie sympathie te winnen bij haar doelgroepen door zichzelf als medeslachtoffer neer te zetten.

In het experiment werd gekeken welke van deze strategieën het best presteerde in de verschillende typen crises. Omdat de lijdstrategie eigenlijk alleen van toepassing is bij externe oorzaak en lage beheersbaarheid en om redenen van een niet te complex experiment is ervoor gekozen om deze responsestrategie en dit type crisis niet mee te nemen in het experiment. Daarmee is het een drie (typen crises) bij vier (responsestrategieën) experiment.

Er waren een viertal experimentele groepen (A,B,C en D), die ieder drie crises (crisis 1 t/m 3) kregen voorgelegd. De verschillen in scores tussen de typen crises werden dus binnen één en dezelfde respondent gemeten (within-respondents). Maar de groepen kregen onderling verschillende response strategieën (responsestrategie I, II, III en IV) gepresenteerd. De effecten van de response strategieën werden dus tussen groepen respondenten gemeten (between-respondents). Iedere groep kreeg dezelfde drie crises voorgelegd, maar gecombineerd met een andere responsestrategie.

Er werd gebruik gemaakt van vier daadwerkelijke journalisten om de krantenberichten zo echt mogelijk te laten zijn. Deze journalisten kregen beknopte informatie van de onderzoeker als ware het een persbericht van de fabrikant. Op basis daarvan schreven zij een krantenbericht. Deze werden in een sjabloon gegoten zodat zij exact leken op een krantenbericht (zie voorbeeld p. 197). Tabel 3 vat de experimentele opzet samen.

Voorbeeld van een krantenbericht over een probleem bij erwtensoeep (interne oorzaak en hoge beheersbaarheid) gecombineerd met een distantiestrategie. Het bericht is geanonimiseerd.

<ORGANISATIE>-soep gevaarlijk voor gezondheid

Van onze verslaggever

ROTTERDAM

De Stevige Erwtensoeep van <ORGANISATIE> is vervuild met het besmettingsmiddel ptoluueensulfanamide. De soep bevat vijf keer zoveel dan wettelijk is toegestaan. Daarmee is het een gevaar voor de gezondheid. De Keuringsdienst van Waren heeft het bedrijf opdracht gegeven het product van de markt te halen.

<ORGANISATIE> gebruikt de stof ptoluueensulfanamide om de houdbaarheid van de soep te vergroten. AKZO-Nobel chemie, producent van het ontsmettingsmiddel meldt dat de stof inderdaad schadelijk is. <ORGANISATIE> claimt dat de stof geen gevaar oplevert.

De Keuringsdienst waarschuwt zwangere vrouwen, kinderen en ouderen het product voorlopig niet te gebruiken.

Uit onderzoek blijkt dat vooral zwakkeren last hebben van de stof. Bij grote hoeveelheden krijgen zij diarree en misselijkheid.

<ORGANISATIE> onkundig

Van onze verslaggever

ROTTERDAM

<ORGANISATIE> wist niet dat het ontsmettingsmiddel ptoluueensulfamide schadelijk is voor de gezondheid. Dat verklaarde een woordvoerder van <ORGANISATIE> gisteren tijdens een persconferentie.

Het management van <ORGANISATIE> bood de Nederlandse consument haar verontschuldigen aan. De woordvoerder benadrukte dat als het bedrijf op de hoogte was geweest van de schadelijke effecten, het bedrijf naar alternatieve oplossingen had gezocht.

De stof ptoluueensulfanamide wordt door <ORGANISATIE> gebruikt om de houdbaarheid van de soep te vergroten. De stof is in grote hoeveelheden schadelijk voor de gezondheid.

INHOUD

ECONOMIE	2
PERSOONLIJK	6
JAN BLOKKER	7
FAMILIEBERICHTEN	10
FORUM	11
COMMENTAAR	11
GEACHTEDIRECTIE	11
KUNST	14
SPORT	15
BEURS	17

HET WEER

Verwachtingen tot vanavond

In het zuiden overwegend bewolkt en mogelijk wat regen. Later op de dag ook zon. Middagtemperatuur ongeveer 6 graden. Noord-oosten wind, matig, kracht 3 tot 4, langs de kust toenemend tot krachtig, 6.

Uitgebreid weeroverzicht op pagina 19

Tabel 3. Design van het experiment.

<i>Type crisis</i>	<i>Crisis response strategie</i>	<i>Journalist</i>	<i>Groep respondenten min. 20 per groep</i>
Crisis 1: Extern / hoge beheersbaarheid	I. Verwijsstrategie	journalist 1	A
Crisis 1: Extern / hoge beheersbaarheid	II. Distantiestrategie	journalist 2	B
Crisis 1: Extern / hoge beheersbaarheid	III. Herinneringsstrategie	journalist 3	C
Crisis 1: Extern / hoge beheersbaarheid	IV. Boetekleedstrategie	journalist 4	D
Crisis 2: Intern / lage beheersbaarheid	I. Verwijsstrategie	journalist 2	B
Crisis 2: Intern / lage beheersbaarheid	II. Distantiestrategie	journalist 3	C
Crisis 2: Intern / lage beheersbaarheid	III. Herinneringsstrategie	journalist 4	D
Crisis 2: Intern / lage beheersbaarheid	IV. Boetekleedstrategie	journalist 1	A
Crisis 3: Intern / hoge beheersbaarheid	I. Verwijsstrategie	journalist 4	D
Crisis 3: Intern / hoge beheersbaarheid	II. Distantiestrategie	journalist 1	A
Crisis 3: Intern / hoge beheersbaarheid	III. Herinneringsstrategie	journalist 2	B
Crisis 3: Intern / hoge beheersbaarheid	IV. Boetekleedstrategie	journalist 3	C

De volgende afhankelijke variabelen zijn in het onderzoek meegenomen.

- Evaluatie van de reactie van de onderneming (“Wat vond u van de reactie van de onderneming op het incident”, 1 item)
- Ondernemingsimago voor de response (1 item)
- Ondernemingsimago na de response (1 item)
- Emotie (zoals angst, boosheid en frustratie) (4 items, Cronbach’s $\alpha = .92$)
- Schuld (4 items, Cronbach’s $\alpha = .81$)
- Koopintentie (negatief geformuleerd, 1 item)

In tabel 4 staan de resultaten vermeld. De rijen vertegenwoordigen de verschillende typen crises van de drie krantenartikelen die de respondenten te lezen hebben gekregen. In de kolommen staan de verschillende onafhankelijke variabelen. Iedere cel geeft de gemiddelde score van het betreffende item of de schaal weer. De onderste rij geeft het significantieniveau aan van het overall effect van het type crisis.

Tabel 4. Type incident effecten op afhankelijke variabelen (gemiddelden).

	<i>Evaluatie van de reactie (hoge scores zijn positiever)</i>	<i>Imago van de organisatie (hoge scores zijn positiever)</i>	<i>Emotie (hoge scores zijn negatiever)</i>	<i>De mate waarin de schuld wordt gegeven aan de organisatie (hoge scores zijn negatiever)</i>	<i>De mate waarin men aangeeft het merk niet meer te zullen kopen (hoge scores zijn negatiever)</i>
Crisis 1: Extern / Hoge beheersbaarheid	5.45	5.19	10.49	14.37	2.16
Crisis 2: Intern / Lage beheersbaarheid	4.91	4.72	11.88*	16.39	3.01
Crisis 3: Intern / Hoge beheersbaarheid	3.12	3.22	18.26	18.99	4.22
<i>Significantie van het type crisis effect (between-respondents)</i>	.000	0.00	.000	.000	.000

* contrast niet significant

Er blijkt een sterk effect voor het type incident zijn. Een incident met een interne oorzaak en hoge beheersbaarheid blijkt de meest negatieve gevolgen te hebben voor de organisatie. Een incident met een interne oorzaak en lage beheersbaarheid heeft minder negatieve gevolgen en een externe oorzaak en hoge beheersbaarheid heeft de minste negatieve gevolgen. Deze conclusie geldt voor alle afhankelijke variabelen. Alle overall effecten¹ zijn significant op 5% niveau en alle contrasten zijn significant op 5% niveau, behalve het verschil tussen crisis 1 en crisis 2 met emotie als onafhankelijke variabele.

Met betrekking tot de responsestrategieën zijn de resultaten minder eenduidig, maar zij bevestigen wel dat de attributie van een crisis in sterke mate bepalend is voor het succes van een responsestrategie. Of sterker geformuleerd, dat de responsestrategie dient te worden afgestemd op de attributie van de crisis door de doelgroepen.

In tabel 5 staan de resultaten van de ANOVA analyses samengevat. In de tabel staat de rangorde van de gemiddelde score van de afhankelijke variabele aangegeven. 1 betekent de meest positieve score en 4 de meest negatieve score. Positief betekent in dit geval: het minst schadelijk voor de organisatie. Indien het contrast niet significant was, staan de scores op een gedeelde plaats. Dus als er in de tabel bij twee responsestrategieën vermeld staat 2/3, dan betekent dat de contrast tussen de twee responsestrategieën niet significant was en er dus sprake is van een gedeelde plaats. Alle effecten zijn significant op een 5% niveau, tenzij anders aangegeven.

¹ De analyses werden gedaan met een repeated measures ANOVA, omdat de type incident effecten within-respondents zijn gemeten.

Tabel 5. Relatieve positie van de gemiddelde score gebaseerd op contrasten en de LSD post-hoc test.

Situatie	Afhankelijke variabelen	Strategieën			
		Verwijsstrategie	Distantieerstrategie	Herinneringsstrategie	Boetekleedstrategie
Extern Hoge beheersbaarheid	Reactie	4	2/3	2/3	1
	Imago	4	1/2/3	1/2/3	1/2/3
	Emotie	4	1/2/3	1/2/3	1/2/3
	Schuld	2/3/4	1	2/3/4	2/3/4
	Koopintentie	3/4	1/2	3/4	1/2
Intern Lage beheersbaarheid	Reactie	4	3	2	1
	Imago	4	1/2/3	1/2/3	1/2/3
	Emotie	4	1/2/3	1/2/3	1/2/3
	Schuld**	1/2	1/2	3/4	3/4
	Koopintentie	3/4	3/4	1/2	1/2
Intern Hoge beheersbaarheid	Reactie*				
	Imago*				
	Emotie*				
	Schuld*				
	Koopintentie*				

1 = minst schadelijk voor de onderneming en 4 = meest schadelijk

* geen significant effect; ** significant op 10% niveau

In het geval van een interne oorzaak en hoge beheersbaarheid is geen van de responsestrategieën in staat om de effecten voor de organisatie te verminderen. Bij de andere incidenten blijkt steeds de boetekleedstrategie het best te presteren. De verwijsstrategie wordt in geen enkel geval geaccepteerd, ook niet als de oorzaak buiten de organisatie ligt (type 4). In het geval van een externe oorzaak en hoge beheersbaarheid blijkt de distantieerstrategie een goed alternatief te kunnen zijn voor de boetekleedstrategie. Wanneer men geconfronteerd met een incident met een interne oorzaak en lage beheersbaarheid blijkt de herinneringsstrategie een goed alternatief te zijn.

6. CONCLUSIE

Voor managers blijkt de attributie van de oorzaak voor een bepaald incident een belangrijke factor te zijn waar rekening mee gehouden dient te worden, bij het reageren op een crisis. Het onderzoeken van de percepties van doelgroepen en vooral de manier waarop zij de oorzaak toewijzen aan de organisatie of de omgeving, is dan ook belangrijk wanneer men geconfronteerd wordt met crisis. Daarbij is het ook essentieel dat er over de oorzaak van het incident wordt gecommuniceerd. Als de oorzaak niet bekend is, dient alles in staat te worden gesteld om dit te achterhalen. Speculeren of gissen over de oorzaak is iets wat men nooit en te nimmer moet doen.

Dit betekent dat de boodschap, wat je zegt en voornamelijk op welke manier de organisatie met de schuldvraag omgaat, een grote invloed heeft op de schade voor de onderneming. Het blijkt dat de perceptie van de consument beïnvloed kan worden door de juiste boodschap de wereld in te sturen. Omdat in het experiment met een viertal journalisten is gewerkt, als tussenstap in het experiment, is voorzichtig de conclusie te trekken dat als de pers op een goede manier voorgelicht wordt, de boodschap goed wordt overgenomen en er een redelijk homogeen en, gegeven de situatie, een zo positief mogelijke beeldvorming ontstaat.

De vraag wat nu de *juiste* boodschap is, kan alleen beantwoord worden door middel van marktonderzoek. Wil een organisatie in de loop van de crisis doormiddel van een crisis-responsestrategie proberen de schade zoveel mogelijk te beperken, dan dient deze reactie meer gebaseerd te zijn op de perceptie van de klanten en van andere doelgroepen, dan op de eigen perceptie. Vaak wordt vanuit een eigen gevoel van machteloosheid en onschuld gecommuniceerd. Bij heel veel crises is men geneigd om de oorzaak buiten de organisatie te leggen en hierdoor de rol van de organisatie en de schade te beperken. Uit dit onderzoek blijkt dat dit nooit door de consument gepikt wordt. Automatisch grijpen naar de boetekleedstrategie is ook niet altijd verstandig. Marktonderzoek naar de perceptie van de klanten levert informatie op over welke strategieën geschikt zijn voor de organisatie. Volgens onderstaand schema kan dan, met aanvullende informatie vanuit het marktonderzoek, een passende responsestrategie geformuleerd worden. De aanvullende informatie bestaat dan bijvoorbeeld uit de vragen waarmee consumenten en andere stakeholders worstelen, niet aan de crisis gerelateerde imago-kenmerken en specifieke kenmerken van de crisis.

Tabel 6. Welke responsestrategie is wanneer geschikt?

Crisis responsestrategie →	<i>Verwijsstrategie</i>	<i>Distantieerstrategie</i>	<i>Herinneringsstrategie</i>	<i>Boetekleedstrategie</i>	<i>Lijdstrategie</i>
Crisis type 1: Extern / Hoge beheersbaarheid		<i>Tweede keus</i>		<i>Eerste keus</i>	
Crisis type 2: Intern / Lage beheersbaarheid			<i>Tweede keus</i>	<i>Eerste keus</i>	
Crisis type 3: Intern / Hoge beheersbaarheid	<i>Geen verschil tussen responsestrategieën</i>				
Crisis type 4: Extern / Lage beheersbaarheid					<i>Eerste keus</i>

Op basis van tabel 6 komt de vraag naar voren wanneer en waarom men zou afwijken van de boetekleedstrategie? In drie van de vier typen crisis presteert deze strategie het best en in het geval van een interne oorzaak en hoge beheersbaarheid lijkt de boetekleedstrategie ook het meest op zijn plaats. De perceptie van de doelgroepen en de schade van het imago is echter maar één invalshoek waar de schuldvraag een belangrijke rol speelt. Het verhaal rond schadevergoedingen en wettelijke aansprakelijkheid is een andere invalshoek. Er kunnen juridische argumenten zijn om een andere responsstrategie te kiezen. Wanneer een onderneming de boetekleedstrategie kiest is het veel moeilijker om, bijvoorbeeld in het geval van een externe oorzaak en beheersbaarheid, schadevergoeding te eisen van bijvoorbeeld een toeleverancier waar de oorzaak lag. Maar ook percepties binnen de organisatie kunnen reden zijn om een alternatieve strategie te kiezen. Het is moeilijk om een boetekleedstrategie intern te verkopen wanneer de percepties intern zodanig zijn dat de organisatie eigenlijk niets verweten kan worden.

De conclusie is dat marktonderzoek een vast onderdeel van de crisis aanpak moeten zijn. In de crisisvoorbereiding dient ook de logistieke planning van het uitvoeren van een marktonderzoek te worden meegenomen. Het is belangrijk om een goed beeld te hebben van de initiële attributie van de oorzaak van de crisis, omdat dat een goed idee geeft hoe de inhoud van de boodschap zou moeten zijn in de tweede fase van een crisis.

LITERATUUR

- Berge, D. ten, *The first 24 hours*, Basil Blackwell, 1988
- Birch, J., New Factors in Crisis Planning and Response, *Public Relations Quarterly*, Spring 1994, pp. 31-34
- Bosch, F.A.J. van den, Riel, C.B.M. van, Buffering- versus bridging-strategieën en het management van strategische issues, *Bedrijfskunde*, Vol. 68, No. 5, 1996, pp. 44-50
- Center, A.H., *Public Relations Practices: Managerial Case Studies and Problems*, Prentice Hall, U.S.A.,
- Coombs, W.T., Choosing the right words. The development of guidelines for the selection of the 'appropriate' crisis-response strategies, *Management Communication Quarterly*, Vol. 8, No. 4, May 1995, pp. 447-476
- Davidson, W.N., Worrell, D.L., Research notes and communications: The effect of product recall announcements on shareholder wealth, *Strategic Management Journal*, Vol. 3, 1992, pp. 467-473
- Fincham, F.D. & J.M. Jaspars, Attribution of responsibility: from man the scientist to man as lawyer, *Advances in experimental social psychology*, Vol. 13, 1980, pp. 81-138
- Fishbein, M. & I. Ajzen, Attribution of responsibility, *Journal of experimental social psychology*, Vol. 9, 1973, pp. 148-153
- Garber, J., Seliman, M.E.P., *Human Helplessness: Theory and applications*, Academic Press, New York, 1980
- Heider, F., *The Psychology of Interpersonal Relations*, John Wiley & Sons, New York, 1959
- Ice, R., Corporate Publics and Rhetorical strategies: The case of Union Carbide's Bhopal crisis, *Management Communication Quarterly*, Vol. 4, No. 3, February 1991, pp. 341-362
- Jaspars, J., Fincham, F.D., Hewstone, M., *Attribution Theory and Research: Conceptual, Developmental and Social Dimensions*, Academic Press, London, 1983
- Kabak, I.W., Siomkos, G.J., Monitoring recovery after a product harm crisis, *Industrial Management*, May/June 1992, pp. 11-12

- Kelley, H.H., Attribution Theory in Social Psychology, *In Nebraska Symposium on Motivation*, ed. Levine, D., 1967, pp. 192-238
- Kelley, H.H. and Michela, J. L., Attribution theory and research, *Annual Review Psychology*, 1980, pp. 457-501
- Mahoney, P.F., It's an emergency – Do you have a plan?, *Management Review*, January 1993, pp. 45-48
- Marcus, A.A., Goodman, R.S., Victims and shareholders: The dilemmas of presenting corporate policy during a crisis, *Academy of Management Journal*, Vol. 34, No. 2, 1991, pp. 281-305
- Mitroff, I.I., Kilmann, R.H., *Corporate Tragedies*, Praeger, New York, U.S.A., 1984
- Pauchant, T.C., Mitroff, I.I., *Transforming the Crisis-Prone Organization*, Jossey-Bass Publishers, San Francisco,
- Pearson, C.M., Mitroff, I.I., From crisis prone to crisis prepared: a framework for crisis management, *Academy of Management Executive*, Vol. 7, No. 1, 1993, pp. 48-59
- Raaij, W.F. van, Causal attributions in economic behavior, *In Economic Psychology: Intersections in theory and application*. eds. Alan J., MacFayden, H.W., 1986, pp. 353-379
- Raaij, W.F. van, Creëer een gunstig beeld van uw bedrijf, *Harvard Holland Review*, No. 10, Spring 1987, pp. 14-20
- Riezebos, R., Negatieve publiciteit rond merken: Van anticipatie tot wederopbouw van het merk, *In Jaarboek '95-'96 van de Nederlandse Vereniging van Marktonderzoekers*, 1995, pp. 195-229
- Rossiter, J., Percy, L., *Advertising Communications and Promotion Management*, 1995
- Shaver, K.G., *The attribution of blame: causality, responsibility and blameworthiness*, Springer-Verlag, New York, 1985
- Siomkos, G., Shrivastava, P., Responding to product liability crises, *Long Range Planning*, Vol. 26, No. 5, 1993, pp. 72-79
- Sturges, D.L., Communicating through crisis: A strategy for organizational survival, *Management Communication Quarterly*, Vol. 7, No. 3, February 1994, pp. 297-316
- Sturges, D.L., Carrell, B.J., Newsom, D.A., Barrera, M., Crisis communication management: The public opinion node and its relationship to environmental nimbus, *SAM Advanced Management Journal*, Summer 1991, pp. 22-27
- Tybout, A.M., Calder, B.J., Sternthal, B., Using information processing theory to design marketing strategies, *Journal of Marketing Research*, Vol. 18, February 1981, pp. 73-79
- Weinberger, M.G., Romeo, J.B., The impact of negative product news, *Business Horizons*, January-February 1989, pp. 44-50
- Wolfe, A., The corporate apology, *Business Horizons*, March-April 1990, pp. 10-14

12. Wat is het? Een motor, een auto of een eh...?

Consumenten categorisatie van hybride producten

K.P.N. MOREL¹

SAMENVATTING

Een productontwikkelingsstrategie die de laatste jaren aan populariteit heeft gewonnen is het ontwerpen van hybride producten. Twee (of meer) bestaande producten (de bronproducten) worden gecombineerd tot één nieuw product. Hybride producten vormen een probleem voor consumenten in de zin dat ze lastig te categoriseren zijn; het is niet duidelijk tot welke categorie ze behoren. Daardoor bestaat het gevaar dat consumenten nieuwe hybride producten minder makkelijk of zelfs helemaal niet zullen accepteren.

Voor bedrijven is het daarom belangrijk van te voren te kunnen voorspellen in welke categorie hun nieuwe hybride product geplaatst zal worden. De bestaande methoden en maten om categorielidmaatschap te bepalen voldoen echter niet of onvoldoende in het geval van hybride producten. De in de literatuur meest gangbare van deze maten, (proto)typicaliteit, wordt in dit onderzoek gebruikt om categorielidmaatschap te voorspellen en te meten. De resultaten maken duidelijk dat typicaliteit geschikt is om het categorielidmaatschap van hybride producten te voorspellen noch te meten. Aanvullende maten en methoden zijn nodig. Enkele voorbeelden hiervan worden gegeven.

1. HET ONTWIKKELEN VAN HYBRIDE PRODUCTEN ALS STRATEGIE VOOR PRODUCTVERNIEUWING

Om te overleven in een sterk competitieve markt brengen bedrijven voortdurend nieuwe producten op de markt die (nog) beter dan bestaande producten inspelen op de behoeften en wensen van consumenten. De geijkte manier van productvernieuwing is het toevoegen van 'features', dat wil zeggen het toevoegen van een extra functie of productvoordeel aan een bestaand product. Het toevoegen van nieuwe productattributen is in veel gevallen een uiterst succesvolle strategie gebleken om concurrentievoordeel te bewerkstelligen (Van Raaij, Antonides, Oppedijk van Veen en Schoormans 1999). Consumenten prefereren producten met nieuwe features en voor fabrikanten bieden features de mogelijkheid om hun producten te onderscheiden van die van de concurrentie (Thölke 1998).

In het verlengde van de nieuwe feature strategie, nemen fabrikanten in toenemende mate hun toevlucht tot een radicalere wijze van productvernieuwing, namelijk het toevoegen van complete producten aan bestaande producten. Voorbeelden van dergelijke nieuwe productcombinaties, die hier hybride producten zullen worden genoemd, zijn

de televisie-videorecorder, combi-oven (magnetron met heteluchtoven), fax-printer-kopieerapparaat, WAP-telefoon, message-watch, camera-horloge, internetmagnetron, en autonavigatiesysteem. De toename van hybride producten kan hoofdzakelijk worden toegeschreven aan de stormachtige ontwikkelingen op het gebied van computertechnologie en mobiele communicatie. De miniaturisering van informatie technologie heeft het mogelijk gemaakt elektronische functies fysiek te integreren in bestaande (en nieuwe) apparaten waar dat voorheen wegens ruimtegebrek onmogelijk was.

Het introduceren van hybride producten lijkt voordelen te bieden voor zowel fabrikanten als consumenten. In vergelijking met het op de markt brengen van producten met nieuwe features (line extensions), bieden hybride producten fabrikanten de mogelijkheid voor hen nieuwe markten te penetreren of zelfs compleet nieuwe markten te ontwikkelen (brand extensions). Nieuwe markten worden in het algemeen gekenmerkt door een hoog groeipotentieel, ze kunnen vroeg betreden worden ('early entrance') en concurrentie is afwezig of zeer beperkt. Voor consumenten zijn hybride producten aantrekkelijk omdat ze kosten reduceren en efficiëntie verhogen (Mayer 1996): hybride producten besparen ruimte en geld omdat slechts één in plaats van twee, drie of meer producten gekocht hoeft te worden.

Hoewel het ontwikkelen van hybride producten een aantrekkelijke strategie lijkt, schuilt er een adder onder het gras. Terwijl het toevoegen van nieuwe features aan een reeds bestaand product het categorielidmaatschap in het algemeen ongemoeid laat (een 100 Hertz televisie blijft een televisie), geldt dit niet voor het samenvoegen van bestaande producten tot een nieuwe productcombinatie. Neem bijvoorbeeld het camera-horloge. Is dit een horloge of een camera? En waar koop je zo'n product: in een fotozaak of in een elektronicawinkel? Om een nieuw hybride product te kunnen begrijpen, moeten consumenten eerst weten wat het is. Daartoe dienen ze te bepalen in welke productcategorie het nieuwe hybride product het beste past. Hier zit het probleem. Hoe kunnen consumenten vaststellen wat de best passende categorie is voor een product dat qua eigenschappen evenzeer thuishoort in een categorie als in een andere? Is een televisie-videorecorder een televisie of een videorecorder? Is een WAP-telefoon een telefoon of een internetmachine? Consumenten hebben moeite met het categoriseren van hybride producten, omdat ze geneigd zijn elk product aan slechts één categorie toe te kennen (Hampton 1998). Een product kan eigenlijk niet tegelijkertijd tot twee verschillende categorieën behoren. Omdat hybride producten wel in meer dan één categorie tegelijkertijd passen, kunnen ze gevoelens van onzekerheid en conflict oproepen. Consumenten weten niet goed wat ze van hybride producten moeten maken. Dit is een belangrijk gegeven voor bedrijven, omdat producten waarover consumenten onzeker zijn minder snel door hen geaccepteerd zullen worden (Rogers 1983; Ziamou 1998). En zo wordt een ogenschijnlijk voordeel mogelijk een groot nadeel.

De consequentie hiervan is dat bedrijven *van tevoren* willen weten hoe consumenten een nieuw hybride product zullen categoriseren. Die kennis is onontbeerlijk om een optimale marketingstrategie (o.a. marktsegmentatie, product positionering en reclame) voor het nieuwe hybride product te kunnen vaststellen (Yi en Gray 1996) en, nog belangrijker, om te kunnen communiceren wat het hybride product is en wat de voordelen ervan zijn. Neem de BMW C1 bijvoorbeeld, een combinatie van een auto en een motorfiets. Hoe moet dit voertuig gepositioneerd worden: als een auto met de eigenschappen en afmetingen van een motorfiets, of als een motorfiets met het comfort en de veiligheid van een auto? BMW heeft de C1 gepositioneerd als een 'nieuwe vorm van persoonlijke mobiliteit', maar hoe moeten ze deze positionering communiceren? Wat

nog belangrijker is: zullen consumenten deze positionering accepteren? Nemen ze de C1 waar zoals hij wordt geadverteerd? Zullen ze de C1 vergelijken met andere auto's of met andere motorfietsen, en met welke auto's of motorfietsen? Onderschrijven consumenten het relatieve voordeel dat de C1 beweerd wordt te hebben? De C1 ziet eruit als een scooter (in plaats van een motorfiets) met overkapping. Zal dit uiterlijk consumenten overtuigen van de claims van comfort en veiligheid die BMW maakt? En zal de gemiddelde motorrijder door dit uiterlijk aangesproken worden? Of zal het uiterlijk van de C1 consumenten doen denken aan een invalidenkarretje, waarmee het ook een zekere gelijkenis vertoont? En waar kan de C1 het beste verkocht worden: bij autodealers, bij motorhuizen of zelfs bij fietshandelaren?

Zoals het voorbeeld van de C1 laat zien, vergroot de ambiguïteit waardoor hybride producten gekenmerkt worden de complexiteit van het toch al ingewikkelde en risicovolle proces van nieuwe productintroductions. Als bedrijven er niet in slagen om het categorielidmaatschap van een nieuw hybride product te voorspellen of als ze er niet in slagen om de gewenste positionering aan consumenten over te brengen, is de kans op een flop niet denkbeeldig. Dus, om nieuwe hybride producten succesvol op de markt te brengen moeten bedrijven weten hoe deze producten door consumenten gecategoriseerd zullen worden. Dit artikel rapporteert een onderzoek naar het categorisatieproces van hybride producten en geeft aan op welke wijze categorielidmaatschap van hybride producten gemeten kan worden. De uitkomsten van dit onderzoek zijn niet alleen interessant voor de makers van hybride producten, maar ook voor marktonderzoekers. Mede op basis van het huidige onderzoek kunnen methoden en maten worden ontwikkeld waarmee onderzocht kan worden hoe consumenten nieuwe hybride producten categoriseren. Eerst zal echter op basis van de literatuur over categorisatie beschreven worden hoe consumenten het categorielidmaatschap van hybride producten bepalen.

2. HOE CATEGORISEREN CONSUMENTEN HYBRIDE PRODUCTEN?

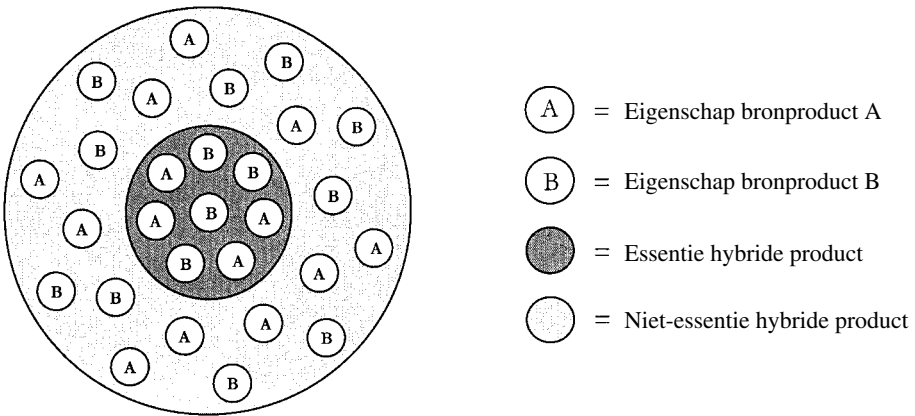
Productcategorisatie bestaat uit het vergelijken van een product met de representatie van dit product in het geheugen. Binnen de literatuur over categorisatie bestaan twee opvattingen over de manier waarop dit vergelijkingsproces plaatsvindt. De dominante, traditionele stroming, de zogenaamde Similarity View, beschouwt categorisatie als het vergelijken van eigenschappen. De eigenschappen van het product worden vergeleken met de eigenschappen van de categorierepresentatie en de mate waarin deze eigenschappen overeenkomen bepaalt of het product al dan niet tot de categorie behoort. Volgens aanhangers van de Theory View, de tweede stroming, is categorisatie meer dan het vergelijken van verzamelingen van eigenschappen. Om te bepalen of een product behoort tot een bepaalde categorie gebruiken mensen zogenaamde naïeve theorieën. Naast informatie over de gelijkenis tussen eigenschappen, omvatten deze theorieën bijvoorbeeld ook informatie over relaties tussen categorieën en over de uitzonderingen en extreme leden binnen categorieën.

Binnen de stroming van de Theory View is één theorie in het bijzonder geschikt om duidelijk te maken waarom hybride producten zo lastig te classificeren zijn. Dit is de theorie van het psychologisch essentialisme (Lakoff 1987, Medin en Ortony 1989, Stevens 2000). Volgens deze theorie geloven mensen dat elk object bestaat uit essentiële eigenschappen ('core properties') en perifere eigenschappen ('peripheral properties').

Essentiële eigenschappen zijn die eigenschappen die het product maken tot wat het is. Zonder die eigenschappen zou het niet dat zijn wat het is, maar iets anders. (Een voorbeeld van een essentiële eigenschap van een televisie is de aanwezigheid van een beeldscherm. Zonder een beeldscherm is een televisie geen televisie meer). Naast essentiële eigenschappen bestaat een product uit perifere eigenschappen. Dit zijn eigenschappen die het product bezit, maar die niet tot de essentie van het product behoren. (De aanwezigheid van stereogeluid op een televisie is zo'n perifere eigenschap. Een televisie zonder stereogeluid blijft een televisie).

Bezien vanuit dit oogpunt bezitten hybride producten een dubbele essentie die bestaat uit de essentiële eigenschappen van elk van de bronproducten (dit zijn de producten waaruit het hybride product is samengesteld). Juist deze dubbele essentie zorgt ervoor dat hybride producten niet eenduidig te classificeren zijn.

Figuur 1 geeft de conceptuele structuur van een hybride product volgens het psychologisch essentialisme weer. Elke productrepresentatie (=concept) bestaat uit een essentie (binnenste donkergrijze cirkel) die de kenmerkende eigenschappen van het concept omvat, en een niet-essentie (buitenste lichtgrijze cirkel) waaronder de perifere eigenschappen vallen. In het geval van een hybride product omvat de essentie zowel essentiële eigenschappen van bronproduct A als van bronproduct B^{2, 3}. Het product heeft dus een dubbele (hybride) essentie: het is zowel A als B. De eigenschappen in het niet-essentiële deel zullen eveneens afkomstig zijn van beide bronproducten, maar bij het vaststellen van het categorielidmaatschap spelen deze eigenschappen geen doorslaggevende rol. Ze zullen hooguit de uiteindelijke classificatie ondersteunen of afzwakken.



Figuur 1. Conceptuele structuur van een hybride product.

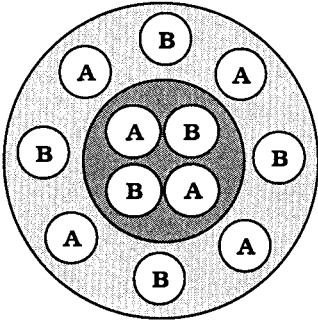
Welke eigenschappen tot de essentie van een concept behoren en welke niet wordt niet duidelijk gemaakt door de theorie van het psychologisch essentialisme. De essentie van een concept bestaat dan ook niet. Wat bestaat is de essentie van een concept zoals die wordt waargenomen door degene die het product categoriseert. Dit betekent dat de essentie van een concept, en dus het categorielidmaatschap, afhankelijk is van de individuele naïeve theorie van de 'categorisator' (degene die categoriseert) over het concept en over de rest van de wereld. Aangezien deze naïeve theorie verschilt van per-

soon tot persoon en kan veranderen in de loop van de tijd, zal ook de *precieze* notie van wat het is dat iets maakt tot wat het is verschillen tussen personen en over tijd⁴. Dit verklaart waarom een en hetzelfde object in de ogen van de ene persoon wel een stoel is en in de ogen van een ander niet. Het neemt echter niet weg dat de *globale* notie van de essentie van een bepaald concept niet of nauwelijks zal verschillen tussen personen. Wanneer verschillende mensen gevraagd wordt te beschrijven wat een stoel is, zullen hun beschrijvingen grotendeels overeenstemmen. Deze overeenstemming vormt de basis voor de productcategorieën die binnen een maatschappij gehanteerd worden.

Het feit dat naïeve theorieën kunnen veranderen over tijd, betekent dat de essentie van een (hybride) product en daarmee het categorielidmaatschap niet vaststaat. Onder invloed van bijvoorbeeld de gebruikservaring die consumenten met een hybride product opdoen, kan het gebeuren dat een product dat in het begin duidelijk hybride was dat na verloop van tijd niet meer is. Iemand die een message-watch koopt zal in het begin waarschijnlijk regelmatig berichten verzenden. Na verloop van tijd, als het nieuwe er af is, blijkt de berichtenfunctie nauwelijks meer gebruikt te worden. Hoewel de message-watch nog steeds beide functies combineert, is het voor de consument een 'gewoon' horloge geworden. De hybride essentie is verloren gegaan.

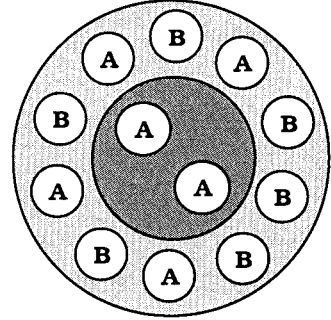
De theorie van het psychologisch essentialisme verschaft een helder beeld van de wijze waarop consumenten bepalen tot welke categorie een hybride product behoort (zie figuur 2). Aangezien consumenten er vanuit gaan dat een product tot die categorie behoort waarvan het de essentiële kenmerken bezit, zullen ze bij het zien van een hybride product in de eerste plaats trachten te bepalen welke kenmerken essentieel zijn voor dit product en in de tweede plaats voor welke categorie of categorieën deze kenmerken essentieel zijn. Aangezien hybride producten zijn ontstaan uit minimaal twee bronproducten, ligt het voor de hand dat consumenten tot de conclusie komen dat de essentiële kenmerken van het hybride product afkomstig zijn uit zowel de ene als de andere broncategorie. In dat geval zullen ze het product dan ook classificeren als het hybride product dat het is (situatie 1 in figuur 2). Het is echter ook denkbaar dat consumenten het hybride product toewijzen aan broncategorie A (B), omdat ze vinden dat enkel essentiële kenmerken van A (B) aanwezig zijn (situaties 2 en 3 in figuur 2). Deze situatie kan zich voordoen als consumenten de essentiële kenmerken van B (A) niet herkennen of niet erkennen. In het eerste geval is de uiteindelijke classificatie het resultaat van onvolledige communicatie (de essentiële eigenschappen van B (A) komen niet duidelijk genoeg naar voren). Een voorbeeld hiervan is dat consumenten de videosleuf in de televisievideo combinatie niet zien en denken dat het een 'gewone' televisie is. Los van dergelijke 'waarnemingsfouten', is in veel gevallen de multifunctionaliteit niet of nauwelijks uit het uiterlijk van het product af te leiden, zoals in het geval van het camera-horloge. In het tweede geval is geen sprake van onvolledige communicatie. De kenmerken van B (A) worden wel degelijk waargenomen, maar zijn in de ogen van de categorisator niet (langer) essentieel voor broncategorie B (A). Ondanks de aanwezigheid van een cameraatje op het camera-horloge, vinden consumenten toch dat het een horloge is, bijvoorbeeld doordat ze de camera vrijwel nooit gebruiken. Gebruikservaring is al eerder aangewezen als een factor die de essentie van een hybride product kan beïnvloeden. Situatie 4 in figuur 2, tot slot, vertegenwoordigt de situatie waarin het hybride product zo weinig associaties met de broncategorieën oproept, dat het noch aan de een noch aan de andere categorie wordt toegewezen. Wat het hybride product dan wel is, is niet bekend.

1. A: aanwezig
B: aanwezig



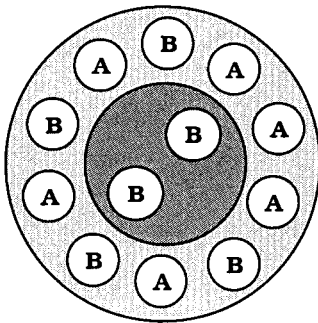
Categorielidmaatschap:
hybride

2. A: aanwezig
B: afwezig



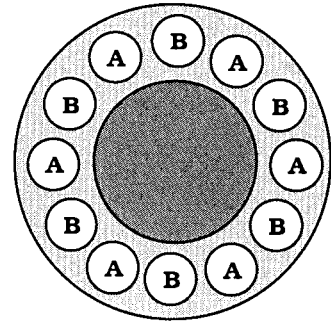
Categorielidmaatschap:
categorie A

3. A: afwezig
B: aanwezig



Categorielidmaatschap:
categorie B

4. A: afwezig
B: afwezig



Categorielidmaatschap:
onbekend

Figuur 2. Vaststellen categorielidmaatschap hybride producten.

Figuur 2 beeldt de vier situaties uit die overeenkomen met de prototypische classificaties van hybride producten die hiervoor besproken zijn. Allerlei varianten op deze vier mogelijkheden zijn denkbaar, waarbij het in figuur 2 veronderstelde evenwicht tussen eigenschappen van A en B afwezig is. Deze mogelijkheden representeren de vele gradaties van categorielidmaatschap die mogelijk zijn binnen elk prototype. Een product dat tot categorie A behoort kan namelijk heel erg A zijn of slechts een beetje A, afhankelijk van de perifere eigenschappen die aanwezig zijn. Ditzelfde geldt voor een product dat als hybride geassocieerd wordt. De mate waarin essentiële eigenschappen van A en B aanwezig zijn kan variëren van hoofdzakelijk A tot hoofdzakelijk B.

3. HET METEN VAN DE CATEGORISATIE VAN HYBRIDE PRODUCTEN

Onder het meten van de categorisatie van hybride producten wordt verstaan het vaststellen van het categorielidmaatschap van hybride producten. Wordt de C1 gezien als een auto, als een motor of als geen van beide? En wat is het dan wel? Aangezien een hybride product eigenschappen in zich verenigt van meer dan één categorie, draait het om de vraag in welke categorie dit product thuishoort. Hoe het proces van categorisatie verloopt is interessant voor zover inzicht in dit proces begrijpelijk maakt in welke categorie hybride producten uiteindelijk geclassificeerd zullen worden.

Een beschouwing van de literatuur over categorisatie binnen de disciplines consumentengedrag en marketing laat echter zien dat de uitkomst van het classificatieproces nauwelijks onderwerp van studie is geweest. Dit categorisatieonderzoek is voor het overgrote deel gebaseerd op het werk van cognitief psychologen zoals Bruner (1977), Rosch (1976) en Barsalou (1983, 1985). Met name naar het begrip prototypicaliteit en de determinanten hiervan is veel onderzoek verricht (o.a. Loken en Ward 1990, Rathneshwar en Shocker 1988, Ward en Loken 1988, Yi en Gray 1996). Verder heeft de aandacht zich gericht op de relatie tussen kennis over een categorie en preferentie voor een product uit die categorie (Coupey, Irwin en Payne 1998, Sujan 1985, Sujan en Bettman 1989) en op de relatie tussen discrepantie met de categorie ('schema incongruity') en de evaluatie van producten uit die categorie (Meyers-Levy en Tybout 1989, Perrachio en Tybout 1996, Stayman, Alden en Smith 1992). Wat opvalt is dat het proces van categorisatie interessanter en belangrijker gevonden wordt dan het resultaat van dit classificatieproces. De meeste studies gaan over de vraag hoe het proces van categorisatie verloopt (analytisch of holistisch?) en over de manier waarop categorieën gepresenteerd worden in het geheugen van consumenten (als verzameling eigenschappen, prototype of meerdere exemplaren?) en niet over de uitkomst, de classificatie. Met andere woorden, categorisatie uitkomsten worden wel gemeten, maar enkel als middel om te bepalen via welk proces het categorielidmaatschap is bepaald. Het idee hierbij is dat verschillende processen leiden tot verschillende uitkomsten. In andere studies gaat de aandacht uit naar de evaluatie van producten waarbij bekendheid met, kennis over of discrepantie met de categorie de onafhankelijke variabele is. Ook hier is het vaststellen van categorielidmaatschap niet het doel (in een aantal gevallen wordt het zelfs helemaal niet vastgesteld), maar het middel om te bepalen in hoeverre het onderzochte product aansluit bij een bepaalde categorie.

Een belangrijke consequentie van de insteek die categorisatie onderzoekers tot dusverre hebben gekozen is dat de gehanteerde onderzoeksmethoden ook gericht zijn op het in kaart brengen van het categorisatieproces in plaats van de uitkomst van dat proces. Twee onderzoeksparadigma's kunnen onderscheiden worden (Basu 1993; Cohen & Basu 1987), namelijk het *category judgment* paradigma en het *verification accuracy* paradigma. De procedure is in beide gevallen gelijk. In de eerste fase leren respondenten een nieuwe categorie aan de hand van een aantal voorbeelden. Ze krijgen bijvoorbeeld reclames te zien van een aantal alcoholische dranken en van een aantal warenhuizen. Vervolgens krijgen ze te horen dat de merken die ze gezien hebben behoren tot één van twee verschillende categorieën, namelijk een categorie "dranken voor na het diner" en een categorie "exclusieve warenhuizen". In de tweede fase krijgen respondenten een aantal nieuwe merken dranken en warenhuizen te zien en moeten ze aangeven tot welke van de geleerde categorieën elk van deze merken behoort. Uit de wijze waarop

respondenten de merken aan de categorieën toewijzen, kan opgemaakt worden of ze een analytisch categorisatieproces hanteren (d.w.z. dat op basis van één of enkele eigenschappen bepaald wordt tot welke categorie een merk behoort) of een holistisch categorisatieproces (waarbij categorielidmaatschap bepaald wordt aan de hand van de algehele globale gelijkheid tussen merken).

Het verschil tussen de twee paradigma's is dat in het *verification accuracy* paradigma respondenten na het classificeren van de merken een aantal vragen over elk van de merken dienen te beantwoorden. Ze dienen bijvoorbeeld aan te geven of een bepaalde eigenschap wel of niet aanwezig was bij een bepaald merk, waarbij deze eigenschap in het ene geval goed past bij het merk en in het andere geval niet goed (bijvoorbeeld de aanwezigheid van ABS bij een Mercedes en bij een Hyundai). De gedachte hierachter is dat mensen die holistisch categoriseren zich zowel de consistente als inconsistente eigenschap correct zullen herinneren, omdat zij de informatie als geheel opslaan inclusief de afwijkende eigenschap. Analytische categorisatoren daarentegen, zullen vaker denken dat de niet-passende eigenschap afwezig was, omdat ze op basis van de aanwezige consistente eigenschappen zullen afleiden dat de inconsistente eigenschap niet aanwezig was. In het geval van de Hyundai met ABS, zullen mensen die holistisch categoriseren zich deze eigenschap herinneren, ondanks het feit dat het wellicht onwaarschijnlijk is dat een Hyundai is uitgerust met ABS. Analytische categorisatoren, die redeneren vanuit hun kennis over het merk, zullen zich deze inconsistente eigenschap zeer waarschijnlijk niet herinneren. Doordat het *verification accuracy* paradigma gebruikt maakt van procesmaten zoals herinnering en informatieverwerking in aanvulling op oordelen over categorielidmaatschap, lijkt deze methode beter geschikt om het categorisatieproces in kaart te brengen dan het *category judgment* paradigma.

Uit de bovenstaande samenvatting van de onderzoeksmethoden die totnogtoe gehanteerd zijn om productcategorisatie in kaart te brengen, blijkt dat er voldoende mogelijkheden zijn om het categorisatieproces in kaart te brengen, maar dat deze niet geschikt zijn om categorielidmaatschap vast te stellen. Viswanathan en Childers (1999, p. 79) komen tot een soortgelijke conclusie naar aanleiding van hun literatuuronderzoek. Volgens deze auteurs zijn de meeste maten van productcategorisatie van een zodanig algemeen niveau dat ze geen inzicht geven in de onderliggende eigenschappen die het categorielidmaatschap bepalen. Zoals gezegd gaat het bij het categoriseren van nieuwe hybride producten juist om het categorielidmaatschap. Om uit te vinden hoe consumenten bepalen in welke categorie een nieuwe hybride product thuis hoort, is een reeks onderzoeken uitgevoerd waarin de categorisatie van hybride producten door consumenten centraal stond. De studie die in de volgende paragraaf gerapporteerd wordt, is één van de onderzoeken uit de reeks. In Morel (2000) wordt het hele onderzoeksproject beschreven.

4. CATEGORIELIDMAATSCHAP VAN HYBRIDE PRODUCTEN: HET EFFECT VAN TYPICALITEIT

In de onderstaande studie is onderzocht of het mogelijk is om het categorielidmaatschap van een hybride product te beïnvloeden door de eigenschappen van de broncategorieën systematisch te variëren. Het doel hierachter was te achterhalen in hoeverre het conceptuele model uit figuur 2 een realistische voorstelling geeft van de relatie tussen de eigenschappen van de broncategorieën en het categorielidmaatschap van het hybride

product. Door middel van experimentele manipulatie van de essentie van de eigenschappen van beide broncategorieën werden de vier situaties uit figuur 2 gecreëerd, waarna gekeken werd of de classificatie van het hybride product overeenkwam met de verwachte classificatie voor elk van de situaties. De mate waarin de eigenschappen van beide broncategorieën essentieel waren voor categorielidmaatschap, werd gemanipuleerd aan de hand van de (proto)typicaliteit van de eigenschappen. Onder (proto)typicaliteit wordt de mate verstaan waarin een eigenschap gedeeld wordt door de leden van een categorie (Smith & Medin 1981). Eigenschappen die door veel leden gedeeld worden zijn de meer centrale of kenmerkende eigenschappen van de categorie (bijvoorbeeld de aanwezigheid van een toetsenbord bij een PC). Producteigenschappen die een hoge typicaliteit voor een bepaalde categorie bezitten zullen tot de essentie van die categorie behoren, terwijl laag-typische eigenschappen daar niet toe zullen behoren. Door de typicaliteit van de eigenschappen van een product zodanig te veranderen dat deze eigenschappen meer of minder typisch zijn voor de bronproducten waaruit het hybride product is opgebouwd, kan de essentie van het hybride product beïnvloed worden en daarmee de classificatie. Op deze wijze kan onderzocht worden van welke eigenschappen de classificatie van hybride producten afhankelijk is en welke eigenschappen dus tot de essentie van een bepaald hybride product behoren en welke niet. Deze informatie is van belang voor bedrijven die willen weten in welke categorie hun nieuwe hybride product door consumenten gestopt zal worden.

De voorspellingen van de studie kunnen als volgt worden samengevat:

H1 Indien een hybride product hoog-typische eigenschappen van zowel het ene als het andere bronproduct bezit, zal het ook als een zelfstandig hybride product geclassificeerd worden (situatie 1 in figuur 2).

H2 Indien hoog-typische eigenschappen van het ene bronproduct aanwezig zijn en laag-typische eigenschappen van het andere bronproduct, zal de hybride geclassificeerd worden in de broncategorie waarvan hoog-typische eigenschappen aanwezig zijn (situaties 2 en 3 in figuur 2).

Als enkel laag-typische eigenschappen aanwezig zijn, is niet duidelijk hoe respondenten het hybride product zullen categoriseren. Hieronder wordt uiteengezet hoe bovenstaande voorspellingen getoetst zijn. In het kader van dit hoofdstuk wordt het onderzoek niet volledig en in detail beschreven, maar alleen de globale opzet en de resultaten die relevant zijn voor het toetsen van de voorspellingen. Zie voor een gedetailleerde beschrijving van het onderzoek Morel (2000).

4.1. Steekproef en experimenteel design

Aan het experiment namen 114 leden (62 mannen en 52 vrouwen, gemiddelde leeftijd van 43 jaar) van het panel van het Product Evaluatie Laboratorium van de Technische Universiteit in Delft deel. De deelnemers werden toegewezen aan één van vijf experimentele groepen. Het design was een incompleet gerandomiseerd 2 x 2 x 2 'between-subjects' design (zie tabel 1).

De afhankelijke variabele was het categorielidmaatschap van het hybride product. De onafhankelijke variabelen waren bekendheid van het hybride product (hoog, laag), typicaliteit van de eigenschappen van broncategorie A (hoog-typisch, laag-typisch, en

Tabel 1. Experimenteel design (N = 114).

Hoge bekendheid				Lage bekendheid			
Laag-typische eigenschappen bron categorie A		Hoog-typische eigenschappen bron categorie A		Laag-typische eigenschappen bron categorie A		Hoog-typische eigenschappen bron categorie A	
Laag-typische eigenschappen bron categorie B	Hoog-typische eigenschappen bron categorie B	Laag-typische eigenschappen bron categorie B	Hoog-typische eigenschappen bron categorie B	Laag-typische eigenschappen bron categorie B	Hoog-typische eigenschappen bron categorie B	Laag-typische eigenschappen bron categorie B	Hoog-typische eigenschappen bron categorie B
n = 23	n = 22	n = 22	n = 23	n = 0	n = 0	n = 0	n = 24

typicaliteit van de eigenschappen van broncategorie B (hoog-typisch, laag-typisch). De experimentele conditie lage bekendheid x hoog-typisch A x hoog-typisch B was toegevoegd om te kunnen controleren voor mogelijke effecten van bekendheid op de classificatie van het hybride product. Deze conditie zal hier verder niet besproken worden.

4.2. Stimulusmateriaal en procedure

Het stimulusmateriaal bestond uit vier verschillende korte productbeschrijvingen van het hybride product televisie-videorecorder combinatie (TV-VR) bestaand uit de broncategorieën televisie (TV) en videorecorder (VR). Elke beschrijving begon met dezelfde introductie, namelijk: "Hieronder staan de eigenschappen van een product dat door bedrijf X op de markt wordt gebracht. Wilt u deze eigenschappen aandachtig doorlezen?". Een lijstje met acht eigenschappen volgde, bestaand uit vier eigenschappen van een televisie en vier eigenschappen van een videorecorder. Deze eigenschappen waren ofwel hoog-typisch ofwel laag-typisch voor de betreffende broncategorie en werden zodanig gecombineerd dat voor elke experimentele conditie één beschrijving verkregen werd. Om te controleren voor mogelijke volgorde effecten (deze bleken na analyse niet te zijn opgetreden), werden de eigenschappen om en om genoemd en de volgorde werd voor de helft van de respondenten omgedraaid. Tot slot dient opgemerkt te worden dat de beschrijvingen identiek waren op alle punten behalve de specifieke eigenschappen die genoemd werden. De hoeveelheid informatie die respondenten te verwerken kregen verschilde niet tussen de condities.

Om de beschrijvingen te kunnen samenstellen, werden eerst zowel hoog- als laag-typische eigenschappen voor de televisie en videorecorder verzameld. Hiertoe schreven 49 respondenten (andere dan in het feitelijke experiment) zoveel mogelijk attributen van elke categorie op als ze konden bedenken. Daarna werden er uit de totale lijst van eigenschappen 15 tot 20 geselecteerd en voorgelegd aan 19 (weer andere) respondenten die elke eigenschap een typicaliteitsscore toekenden. Dit deden ze door voor elke eigenschap aan te geven in hoeverre deze eigenschap paste (zeer goed, redelijk, helemaal niet) bij de categorie televisie of videorecorder. Eigenschappen met het grootste aantal 'past heel goed' kwalificaties kregen het predikaat hoog-typisch. Een eigenschap werd als laag-typisch gekwalificeerd, indien deze volgens de meeste mensen goed tot

Tabel 2. Hoog- en laag-typische eigenschappen voor de televisie en videorecorder die gebruikt zijn in de productbeschrijvingen.

Categorie	Eigenschap (typicaliteit)	Past heel goed	Past redelijk goed	Past niet	Totaal
Televisie					
HT TV	Het product heeft een beeldscherm (hoog)	19	0	0	19
	Op het product kun je de zender kiezen (hoog)	18	0	1	19
	Met het product kun je programma's bekijken (hoog)	18	0	1	19
	Op het product kun je de kanalen programmeren (hoog)	15	4	0	19
LT TV	Het product is zwaar (laag)	4	11	4	19
	Het product bepaalt de indeling van de huiskamer (laag)	6	7	6	19
	Het product heeft ondertiteling (laag)	9	7	3	19
	Het product geeft de huidige zender weer (laag)	9	7	3	19
Video					
HT VR	Er kan een band in het product (hoog)	19	0	0	19
	Het product kan opnemen (hoog)	18	0	1	19
	Je kunt het product programmeren (hoog)	18	1	0	19
	Het product kan vooruitspoelen en terugspoelen (hoog)	18	1	0	19
LT VR	Het product geeft de speelduur aan (laag)	12	6	1	19
	Het product schakelt na bepaalde tijd automatisch over op stand-by (laag)	8	8	3	19
	Het product heeft een groot display met veel informatie (laag)	6	9	4	19
	Er zitten veel knoppen op het product (laag)	3	10	6	19

redelijk goed paste. Tabel 2 geeft de uitkomsten van deze pretest voor de eigenschappen die uiteindelijk gebruikt zijn in de productbeschrijvingen.

In een tweede pretest is vervolgens onderzocht of de uiteindelijke productbeschrijvingen beide beoogde broncategorieën oproepen. Dit was belangrijk, omdat respondenten in alle gevallen, dus ook wanneer de eigenschappen in de beschrijving laag-typisch waren, in de gaten moesten hebben dat het om een hybride product ging. Alle vier de beschrijvingen van de TV-VR moesten dus de categorieën televisie en videorecorder oproepen. De resultaten van deze pretest (hier niet gegeven) lieten zien dat dit inderdaad het geval was.

4.3. Maten van categorielidmaatschap

Het onderzoek was opgezet om de categorisatie van de hybride TV-VR te bestuderen. Deze variabele is op drie manieren gemeten. In eerste instantie schreven respondenten net als in de tweede pretest op aan welke producten (productcategorieën) het product uit de beschrijving hen deed denken. In aanvulling op deze spontane maat voor catego-

rielandmaatschap, gaven respondenten aan wat het product volgens hen was, waarbij ze de keus hadden uit vier opties, namelijk (1) videorecorder; (2) televisie; (3) televisie-videorecorder; of (4) iets anders. Tot slot werd de mate van typicaliteit van het product uit de beschrijving voor elk van de broncategorieën en voor de hybride categorie gemeten aan de hand van de volgende vraag: "Wilt u aangeven in hoeverre het product uit de beschrijving volgens u in de categorie televisies (videorecorders, televisie-videorecorders) past?", waarbij het antwoord kon variëren van 1 (past heel erg goed) tot 7 (past heel erg slecht). Zo werden drie typicaliteitsmaten verkregen, één voor elke broncategorie en één voor de hybride categorie .

4.4. Resultaten

Tabel 3 geeft de resultaten van de spontane categorisatietaak. De tabel geeft het percentage van de respondenten dat de betreffende categorie (televisie, videorecorder, televisie-videorecorder) noemde binnen elke conditie. De cellen waarin relatief hoge frequenties voorspeld werden zijn cursief gedrukt.

In hoeverre komen deze resultaten overeen met de hypothesen? In die condities waarin laag-typische eigenschappen aanwezig waren, moest het percentage respondenten dat de overeenkomstige broncategorie noemde relatief laag zijn. Waar hoog-typische eigenschappen aanwezig waren, werden relatief hoge percentages voor de overeenkomstige broncategorie verwacht. Een hoog percentage respondenten dat de hybride categorie televisie-videorecorder noemde werd voorspeld in de HTTV x HTVR conditie, omdat dan volgens de veronderstelling essentiële eigenschappen van beide bronproducten aanwezig waren. Het patroon van cursief gedrukte percentages geeft aan dat de uitkomsten redelijk tot goed overeenkwamen met de verwachtingen. Wanneer alle eigenschappen laag-typisch waren (conditie 1), werden beide broncategorieën even vaak genoemd. In conditie 2, waarin de eigenschappen van de televisie hoog-typisch waren en die van de videorecorder laag-typisch, werd de televisie categorie vaker wel (59.1) dan niet (40.9) genoemd en de videorecorder categorie vaker niet (68.2) dan wel (31.8). In de derde conditie (LTTV x HTVR) is dit patroon precies omgekeerd, zoals verwacht werd. In de HTTV x HTVR conditie voldeden de bevindingen niet aan de verwachtingen. Het percentage respondenten in deze conditie dat de televisiecategorie noemde is opmerkelijk laag (30.4) waar een hoog percentage verwacht werd. Dit is des te opvallender, aangezien in de andere conditie waarin precies dezelfde hoog-typische televisie-eigenschappen beschreven werden (conditie 2), het percentage dat de televisiecategorie noemde wel hoog was, zelfs bijna twee keer zo hoog (59.1). Blijkbaar maakte het voor respondenten verschil welke eigenschappen van de videorecorder (laag-typisch of hoog-typisch) in de beschrijving stonden. Op zich is dat niet zo verwonderlijk, want respondenten krijgen immers verschillende informatie aangeboden in de verschillende condities. Maar daarmee valt niet te verklaren waarom dit verschil niet optreedt in het geval van de hoog-typische eigenschappen van de videorecorder. Het percentage respondenten dat de videorecorder noemde, was namelijk wel hoog in beide condities (LTTV x HTVR: 77.3; HTTV x HTVR: 69.6). Ondanks het feit dat de typicaliteit van alle eigenschappen van te voren getest was, is het blijkbaar niet vanzelfsprekend dat deze eigenschappen even typisch zijn wanneer ze gecombineerd worden met andere eigenschappen. Het antwoordpatroon voor de hybride categorie televisie-videorecorder laat zien dat, zoals verwacht, alleen in de vierde conditie (HTTV x HTVR) meer dan de helft van de respondenten deze categorie spontaan noemde. In de overige

condities werd het noemen van deze hybride categorie waarschijnlijk belemmerd door de aanwezigheid van hoog-typische eigenschappen van één van beide bronproducten (condities 2 en 3) of door de totale afwezigheid van hoog-typische eigenschappen (conditie 1).

Tabel 3. Spontane categorisatie: percentage respondenten per conditie dat betreffende categorie noemt.¹

	Conditie ²							
	1		2		3		4	
	LT TV LT VR	Nee	HT TV LT VR	Nee	LT TV HT VR	Nee	HT TV HT VR	Nee
Categorie genoemd	Ja	Nee	Ja	Nee	Ja	Nee	Ja	Nee
Televisie	52.2	47.8	59.1	40.9	40.9	59.1	30.4	69.6
Videorecorder	52.2	47.8	31.8	68.2	77.3	22.7	69.6	30.4
Televisie-videorecorder	21.7	78.3	18.2	81.8	13.6	86.4	52.2	47.8
Aantal waarnemingen (n)	23		22		22		23	

¹ Respondenten konden meer dan één categorie noemen.

² LT = laag-typisch; HT = hoog-typisch.

De hypothese dat respondenten in de HTTV x HTVR conditie het product uit de beschrijving vaker als een hybride televisie-videorecorder zouden categoriseren dan respondenten in de overige condities (H1) werd getoetst aan de hand van de resultaten op de gesloten classificatietaak (zie tabel 4), de tweede maat van categorielidmaatschap. Uit tabel 4 valt op te maken dat de frequentieverdeling tegen de verwachting in vrijwel identiek was in de vier condities. Een χ^2 -toets leverde geen significante verschillen op ($p = .166$). Alhoewel het percentage het hoogst was in de HTTV x HTVR conditie, classificeerde de overgrote meerderheid van de respondenten in de andere condities (gemiddeld 77.7 procent) het product uit de beschrijving eveneens als een televisie-videorecorder. Deze observatie ondermijnt ook de tweede hypothese dat de hybride geassocieerd zou worden in de broncategorie waarvan hoog-typische eigenschappen aanwezig waren. Een andere opvallend resultaat is dat in de LTVR condities 4 respondenten dachten dat het beschreven product iets anders was dan een televisie, videorecorder, of een combinatie van deze twee. Wat het product wel was volgens deze personen, is niet bekend.

Tabel 4. Kruistabel experimentele conditie en classificatie.

Conditie	Classificatie				Totaal
	VR (%)	TV (%)	TV-VR (%)	Anders (%)	
LT TV x LT VR	1 (4.3)	1 (4.3)	17 (73.9)	4 (17.4)	23
HT TV x LT VR	0 (0.0)	1 (4.5)	17 (77.3)	4 (18.2)	22
LT TV x HT VR	3 (13.6)	1 (4.5)	18 (81.8)	0 (0.0)	22
HT TV x HT VR	2 (8.7)	0 (0.0)	21 (1.3)	0 (0.0)	23
Totaal	6 (6.7)	3 (3.3)	73 (81.1)	8 (8.9)	90

De resultaten met betrekking tot de derde maat van categorielidmaatschap (de gemiddelde typicaliteit van elke beschrijving voor de televisie-, de videorecorder, en de televisie-videorecorder categorie) zijn gegeven in tabel 5.

Tabel 5. Gemiddelde typicaliteit voor de broncategorieën en de hybride categorie.

Conditie	Televisie	Videorecorder	Televisie-videorecorder
LT TV x LT VR	5.17	3.43	6.09
HT TV x LT VR	5.36	3.09	6.36
LT TV x HT VR	4.73	2.50	6.59
HT TV x HT VR	4.26	3.22	6.13

Uit de gemiddelden valt duidelijk op te maken dat respondenten in elke conditie de productbeschrijving heel goed vonden passen in de televisie-videorecorder categorie (gemiddeld 6.29 op een 7-puntsschaal). Dit resultaat komt overeen met de zeer hoge percentages in de kolom TV-VR uit tabel 4. Op basis van deze resultaten lijkt het erop dat het respondenten voldoende duidelijk was dat de beschrijvingen betrekking hadden op een combinatie van een televisie en videorecorder, maar dat ze minder gevoelig bleken te zijn voor de verschillen tussen de condities. Dit idee wordt ondersteund door de scores met betrekking tot de broncategorieën. Met uitzondering van de HTTV x LTVR conditie, werden hoge scores gevonden waar lage verwacht werden en vice versa. Twee 2 x 2 ANOVA's met als afhankelijke variabelen de typicaliteit voor de televisie categorie en de typicaliteit voor de videorecorder categorie bevestigden dat er geen systematische significante verschillen tussen de condities waren. Indien de experimentele manipulatie geslaagd was, had in beide analyses een significant interactie-effect gevonden moeten worden. Dit was niet het geval. Het enige significante effect was een hoofdeffect van typicaliteit van de videorecorder eigenschappen op de waargenomen typicaliteit voor de televisie categorie. Die beschrijvingen waarin de eigenschappen van de videorecorder laag-typisch waren werden als meer typisch voor de televisie categorie gezien ($M = 5.27$) dan de beschrijvingen met hoog-typische videorecorder eigenschappen ($M = 4.49$). Net als de resultaten op de gesloten classificatietoets wekken deze resultaten de indruk dat de experimentele manipulatie van de typicaliteit van de eigenschappen van de broncategorieën niet voldoende gewerkt heeft. Dit is verrassend, aangezien deze eigenschappen geselecteerd zijn op basis van een vooronderzoek waaruit bleek dat de typicaliteit wel degelijk verschilde.

4.5. Het effect van de typicaliteit van de broncategorieën op de categorisatie van hybride producten

Dit onderzoek had tot doel de mogelijkheid te onderzoeken om het categorielidmaatschap van een hybride product (TV-VR) te beïnvloeden door de essentie van het hybride product systematisch te variëren aan de hand van de typicaliteit van de eigenschappen van de broncategorieën. De resultaten suggereren dat dit slechts ten dele gelukt is. Het aanbieden van verschillende combinaties van hoog- en laag-typische eigenschappen van de televisie en van de videorecorder lijkt wel degelijk invloed te hebben gehad op de manier waarop respondenten tegen de TV-VR aankeken, maar die invloed heeft

niet op alle categorisatiematen tot significante verschillen geleid. Wanneer respondenten gevraagd werd de productcategorieën te noemen waaraan de beschrijving hen deed denken, bleek dat de broncategorie waarvan hoog-typische eigenschappen aanwezig waren, vaker werd genoemd dan de broncategorie waarvan laag-typische eigenschappen in de beschrijving opgenomen waren. Bovendien bleek dat de hybride categorie televisie-videorecorder door maximaal 20 procent van de respondenten genoemd werd in alle condities, behalve in de conditie waarin de eigenschappen van zowel de televisie als videorecorder hoog-typisch waren en waarvoor het percentage ruim 50 procent was. Bij elkaar genomen onderschrijven deze bevindingen beide hypothesen. Wanneer respondenten echter het product uit de beschrijving dienden toe te kennen aan één van vier alternatieven, is het beeld bevestigend nog ontkennend. Zoals voorspeld werd, was het aantal respondenten dat het product toekende aan de televisie-videorecorder categorie het hoogst in de HTTV x HTVR conditie, maar het verschil was niet significant, omdat ook in de overige condities vrijwel iedereen dacht dat het beschreven product een televisie-videorecorder was. Ook bij de resultaten met betrekking tot de typicaliteit van het beschreven product voor de hybride televisie-videorecorder werden nauwelijks verschillen gevonden tussen respondenten uit verschillende condities; ze vonden vrijwel zonder uitzondering dat het beschreven product zeer goed paste in de categorie televisie-videorecorders.

Hoewel er verschillende onderzoekstechnische oorzaken te bedenken zijn waarom respondenten in verschillende condities niet significant verschilden op de gesloten classificatiemaat en op de typicaliteitsmaat (de experimentele manipulatie schoot tekort, de gebruikte categorisatiematen waren niet valide, de hoeveelheid informatie in de beschrijving was te summier), lijkt een meer fundamenteel probleem het onderzoek parten te hebben gespeeld. Het vermoeden dat naar aanleiding van deze studie is ontstaan is dat de relatie tussen de typicaliteit van de eigenschappen van de broncategorieën en de categorisatie van het hybride product niet eenduidig genoeg is. Bij het bepalen van categorielidmaatschap letten consumenten er niet alleen op hoe typisch de eigenschappen van het hybride product zijn voor de ene broncategorie, maar ook hoe atypisch deze eigenschappen zijn voor de andere broncategorie (Hampton 1998). De eigenschap 'heeft een beeldscherm' bijvoorbeeld is typisch voor televisies en atypisch voor videorecorders. De aanwezigheid van deze eigenschap zal daarom classificatie in de televisiecategorie bevorderen en tegelijkertijd classificatie in de videorecordercategorie ontmoedigen. De eigenschap 'is zwart' daarentegen is typisch voor televisies (de meeste televisies zijn zwart), maar ook typisch voor videorecorders die veelal ook zwart zijn. 'Zwartheid' als eigenschap bevordert noch de categorisatie als televisie, noch als videorecorder, omdat het geen onderscheidende eigenschap is. Dat wil zeggen dat zwartheid geen eigenschap is op basis waarvan televisies en videorecorder van elkaar kunnen worden onderscheiden. De aanwezigheid van een beeldscherm is wel een onderscheidende eigenschap, omdat televisies wel en videorecorders niet een beeldscherm hebben.

Het voorgaande betekent niet alleen dat het manipuleren van de typicaliteit van de eigenschappen van de broncategorieën van een hybride product een ontoereikende methode is om het categorielidmaatschap van het hybride product te sturen. Het betekent ook dat kennis over de aanwezigheid van typische eigenschappen van de bronproducten alleen onvoldoende is om het categorielidmaatschap van hybride producten te kunnen voorspellen, omdat niet bekend is in hoeverre de aanwezige eigenschappen atypisch zijn voor het contrasterende bronproduct. Dit laatste punt is zeer relevant voor producenten, aangezien zij *van tevoren* willen weten hoe hun nieuwe hybride product

door consumenten gecategoriseerd zal worden. En daarmee is het ook relevant voor marktonderzoekers die ingeschakeld worden om het categorielidmaatschap van die nieuwe hybride producten te voorspellen. In de laatste paragraaf wordt een maat gesuggereerd als alternatief voor typicaliteit aan de hand waarvan categorielidmaatschap van hybride producten beter beïnvloed en voorspeld kan worden. Daarnaast wordt betoogd dat de in de literatuur dominante maat voor categorielidmaatschap, namelijk prototypicaliteit, minder geschikt is in het geval van hybride producten en wordt een aantal alternatieve manieren voorgesteld waarop het categorielidmaatschap van hybride producten vastgesteld kan worden.

5. AANBEVELINGEN VOOR HET VOORSPELLEN EN METEN VAN CATEGORISATIE VAN HYBRIDE PRODUCTEN

Zoals gezegd, letten consumenten er bij het bepalen van categorielidmaatschap niet alleen op hoe typisch de eigenschappen van het hybride product zijn voor de ene broncategorie, maar ook hoe atypisch deze eigenschappen zijn voor de andere broncategorie. Een maat die zowel rekening houdt met typicaliteit voor de eigen categorie als atypicaliteit voor contrasterende categorieën, is diagnosticiteit. Diagnosticiteit kan opgevat worden als het discriminerend vermogen van een eigenschap. Het begrip is gedefinieerd als de mate waarin een eigenschap leden van een categorie onderscheidt van niet-leden (Yi & Gray 1996). Het onderscheid tussen typicaliteit en diagnosticiteit kan het beste duidelijk gemaakt worden aan de hand van de begrippen 'category validity' en 'cue validity' (Tversky 1977). Category validity heeft betrekking op de waarschijnlijkheid dat een eigenschap aanwezig is, gegeven dat een object tot een bepaalde categorie behoort (bijv. de kans dat een jas een rits heeft). Cue validity verwijst naar de omgekeerde situatie. Het betreft de kans dat iets tot een bepaalde categorie behoort gegeven de aanwezigheid van een bepaalde eigenschap (bijv. de kans dat iets dat een rits heeft een jas is). De mate van category validity bepaalt de mate van typicaliteit. Eigenschappen die een grote kans hebben om tot een bepaalde categorie te behoren, zijn typisch voor die categorie. Om diagnostisch te zijn, moeten typische eigenschappen ook een hoge cue validity hebben. De eigenschap 'heeft een rits' bijvoorbeeld, is een typische eigenschap van zowel een slaapzak als een jas, maar deze eigenschap is niet geschikt om een slaapzak van een jas te onderscheiden, omdat de cue validity laag is. Kortom, typicaliteit is een functie van category validity, terwijl diagnosticiteit een functie is van category validity en cue validity.

Marktonderzoekers die willen voorspellen hoe een nieuw hybride product geclassificeerd zal worden, en met name een antwoord willen geven op de voor fabrikanten belangrijkste vraag of de hybride in een zelfstandige categorie zal komen of niet, dienen dus niet alleen vast te stellen in hoeverre de eigenschappen van het hybride product typisch zijn voor de broncategorieën, maar tevens in hoeverre deze eigenschappen diagnostisch zijn. Diagnosticiteit kan gemeten worden door de gemiddelde category validity en cue validity te berekenen voor de totale set van producteigenschappen waaruit de hybride bestaat (zie Yi en Gray 1996). Uit deze omschrijving blijkt dat diagnosticiteit in tegenstelling tot typicaliteit alleen op attribuutniveau gemeten kan worden. Respondenten kunnen wel aangeven hoe onderscheidend een eigenschap is voor een bepaalde categorie, maar ze kunnen niet zeggen hoe onderscheidend een product is voor een bepaalde categorie. Dit komt doordat categorielidmaatschap discreet is (ter-

wijl *de mate van* categorielidmaatschap gradueel is). Een product behoort wel of niet tot een categorie. De vraag hoe onderscheidend een product voor een categorie is (dit zou een manier kunnen zijn om diagnosticiteit te operationaliseren) is een onzinnige vraag, omdat elk lid van een categorie vanzelfsprekend onderscheidend is aangezien het deel uitmaakt van die categorie en niet van een andere.

Indien onderzoekers categorisatie van hybride producten op productniveau willen bestuderen, bijvoorbeeld omdat het stimulusmateriaal bestaat uit concrete producten en niet uit lijsten van eigenschappen of beschrijvingen, leert deze studie dat het meten van categorielidmaatschap niet eenvoudig is. *Om een betrouwbaar en valide oordeel te krijgen verdient het aanbeveling om een combinatie van categorisatiematen te gebruiken. Het meten van categorielidmaatschap moet in elk geval niet beperkt worden tot het vaststellen van de typicaliteit, zoals dat in het meeste categorisatieonderzoek gebeurt.* Het probleem met het concept prototypicaliteit is dat het weliswaar mogelijk is om te meten of een product een typisch of atypisch lid van een categorie is, maar zolang niet gespecificeerd kan worden waar de categoriegrenzen liggen, dus wanneer een atypisch lid buiten de categorie valt, is het concept niet bruikbaar om categorielidmaatschap te voorspellen. Bovendien vereist het meten van typicaliteit dat de target categorie (de categorie waarmee het product vergeleken wordt) gespecificeerd wordt. Typicaliteit is een relatief begrip. Een product is niet (a)typisch op zich, het is (a)typisch voor een bepaalde categorie. In feite betekent dit dat typicaliteit pas gemeten kan worden als de target categorie(ën) vaststaat(t)(n), dat wil zeggen als het product geassocieerd is. Nu is het meestal geen probleem om te bepalen wat de target categorie is, aangezien deze bepaald wordt door degene die het product op de markt brengt. Maar in het geval van compleet nieuwe producten, zoals hybride producten, waarvoor nog geen categorie bestaat, kan typicaliteit bij gebrek aan een target categorie niet gemeten worden. In zo'n geval zal eerst bepaald moeten worden wat mogelijke target categorieën zijn. Een laatste reden waarom typicaliteit een minder geschikt concept is om categorielidmaatschap van hybride producten te meten is dat typicaliteit normaliter gemeten wordt ten opzichte van een enkele categorie. In het geval van hybride producten volstaat dit niet, aangezien dit type producten prototypisch is voor meerdere categorieën. Prototypicaliteit van een hybride product dient dus bepaald te worden voor elk van de broncategorieën en voor de hybride categorie (als die bestaat).

Als het meten van typicaliteit niet volstaat, *hoe kan het categorielidmaatschap van nieuw hybride producten dan wel gemeten worden?* Een eerste mogelijkheid, die in de huidige studie werd toegepast, is consumenten te laten aangeven wat zij denken dat het nieuwe product is of aan welke andere producten het product hen doet denken. Een andere mogelijkheid bestaat eruit respondenten niet een puur cognitieve classificatietask te laten uitvoeren, maar ze een set met (plaatjes van) concrete producten uit verschillende categorieën te geven en ze te vragen de producten eruit te halen waarop het nieuwe product het meeste lijkt. Deze taak is in feite een gecombineerde vergelijkings- en ordeningstaak, waarbij respondenten in verschillende stappen tot de beste match komen. Als respondenten tevens 'gedwongen' worden een definitieve categorie te kiezen, kan het probleem van de ongedefinieerde categoriegrenzen aangepakt worden. Respondenten kunnen namelijk niet volstaan met een typicaliteitsoordeel, maar ze moeten kiezen of het nieuwe product binnen of buiten een bepaalde categorie valt. Op deze manier wordt ook duidelijk of het nieuwe product in een bestaande categorie geplaatst kan worden of dat een nieuwe categorie gecreëerd moet worden. Eerder is al aangegeven dat het met name in het geval van nieuwe hybride producten, waarbij het

nieuwe product overeenkomsten vertoont met beide bronproducten, van belang is te weten of de hybride in één van de broncategorieën past of een zelfstandige categorie dient te krijgen. Een studie waarin dit onderzocht is en waarbij gebruik is gemaakt van een combinatie van spontane classificatie, ‘gedwongen’ classificatie en typicaliteitsmetingen, is beschreven in Morel (2000, hoofdstuk 6).

Een ander punt dat dit artikel wil benadrukken is *dat de geschetste problematiek rondom hybride producten zich niet beperkt tot dit type producten*. Hybride producten zijn op te vatten als het type complexe stimuli waarmee consumenten tegenwoordig te maken krijgen. In plaats van relatief eenduidige ééndimensionale functionele producten zijn de huidige producten totaalproducten geworden, een complete set van functies, services, waarden en emoties. Denk bijvoorbeeld aan de product-service combinaties die ontstaan als gevolg van het toevoegen van extra waarde aan producten door het leveren van service. Mobiele telefoons zijn nauwelijks meer los te koop, maar enkel in combinatie met een abonnement of prepaid kaart. Computers worden verkocht inclusief software en technische ondersteuning. Boeking voor een bungalowpark of vakantieoord omvatten transport, hotel, eten, excursies en avondjes uit. Maar ook de toegenomen kennis en bewustheid bij consumenten zorgen voor een complexere betekenisstructuur rond producten. Naast functie, vorm en merk, wordt het ook belangrijk in welk land het product gemaakt is, welke materialen gebruikt zijn, hoe de producent omgaat met ethische vraagstukken, enzovoort. Door al deze ontwikkelingen wordt de vraag naar wat een product nou precies is, tot welke categorie het behoort, relevant in een veel bredere context dan die van hybride producten. Inzicht in de wijze waarop consumenten complexe producten (waaronder hybride producten) categoriseren en hoe deze categorisatie gemeten kan worden draagt dus bij aan een dieper begrip van consumentenreacties op een nieuwe generatie producten.

In het verlengde van het voorgaande punt rijst de vraag *in hoeverre het onderzoek naar hybriditeit vertaalbaar is naar andere aandachtsgebieden, zoals bijvoorbeeld reclame*. Een vorm van hybriditeit binnen reclame betreft de relatie tussen de reclameboodschap en de productcategorie waarop de boodschap betrekking heeft. Reclames maken in toenemende mate gebruik van dezelfde thema’s (seksuele aantrekkingskracht, vrijheid, genieten van het leven) terwijl het om totaal verschillende producten gaat (verzekeringen, auto’s, bier, deodorant, drank, of zelfs snoep). Het zou interessant zijn om uit te zoeken in hoeverre het hierboven geschetste categorisatie-onderzoek over hybride producten van toepassing is op andere objecten dan producten of diensten, zoals bijvoorbeeld reclame.

NOTEN

1. De auteur bedankt Jan Schoormans en Theo Poiesz voor hun commentaar op dit artikel.
2. Met het oog op de eenvoud wordt aangenomen dat een hybride product bestaat uit twee bronproducten. Hoewel in werkelijkheid hybride producten bestaand uit twee bronproducten het vaakst voorkomen, bestaan ook hybride combinaties van meer dan twee bronproducten.
3. Naast eigenschappen van de bronproducten A en B, bestaat de mogelijkheid dat de essentie eigenschappen bevat die uniek zijn voor de hybride combinatie (zie Morel 2000). Deze eigenschappen worden hier weggelaten om de uiteenzetting niet onnodig complex te maken.
4. Dit verklaart waarom de speurtocht naar de noodzakelijke en voldoende eigenschappen van een categorie zoals die werd uitgevoerd binnen de traditie van de “Classical View” (zie bijvoorbeeld Smith en Medin 1981) bij voorbaat vruchteloos is.

LITERATUUR

- Barsalou, Lawrence W. (1983), "Ad Hoc Categories", *Memory and Cognition*, Vol. 11, 211-227.
- Barsalou, Lawrence, W. (1985), "Ideals, Central Tendency, and Frequency of Instantiation as Determinants of Graded Structure in Categories", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 11 (4), 629-654.
- Basu, Kunal (1993), "Consumers' Categorization Processes: An Examination With Two Alternative Methodological Paradigms", *Journal of Consumer Psychology*, Vol. 2 (2), 97-121.
- Bruner, Jerome S., Jacqueline J. Goodnow, George A. Austin (1977), *A Study of Thinking* (reprint of the 7th printing of the 1956 edition), Huntington (NY): Robert E. Krieger Publishing Company.
- Cohen, Joel B., and Kunal Basu (1987), "Alternative Models of Categorization: Toward a Contingent Processing Framework", *Journal of Consumer Research*, Vol. 13, 455-472.
- Coupey, Eloïse, Julie R. Irwin, and John W. Payne (1998), "Product Category Familiarity and Preference Construction", *Journal of Consumer Research*, Vol. 24, 459-467.
- Hampton, James A. (1998), "Similarity-based Categorization and Fuzziness of Natural Categories", *Cognition*, Vol. 65, 137-165.
- Lakoff, George (1987), "Cognitive Models and Prototype Theory", in Ulric Neisser (Ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, New York, NY: Cambridge University Press, 63-100.
- Loken, Barbara, and James Ward (1990), "Alternative Approaches to Understanding the Determinants of Typicality", *Journal of Consumer Research*, Vol. 17, 111-126.
- Mayer, Barbara (1996), "Multifunction Communications Gear Foreseen as Major Electronics Trend", *Hudson Valley Business Journal*, Vol. 6, 22.
- Medin, Douglas, and Andrew Ortony (1989), "Psychological Essentialism", in Stella Vosniadou, and Andrew Ortony (Eds.), *Similarity and Analogical Reasoning*, New York, NY: Cambridge University Press, 179-195
- Meyers-Levy, Joan, and Alice M. Tybout (1989), "Schema Congruity as a Basis for Product Evaluation", *Journal of Consumer Research*, Vol. 16, 39-54.
- Morel, K.P.N. (2000), *Consumers' Reactions to Ambiguous Product Information: The Categorization of Hybrid Products*, Delft: Eburon.
- Peracchio, Laura A., and Alice M. Tybout (1996), "The Moderating Role of Prior Knowledge in Schema-Based Product Evaluation", *Journal of Consumer Research*, Vol. 23, 177-192.
- Raaij, W.F. van, G. Antonides, W.M Oppedijk van Veen, en J.P.L. Schoormans (1999), *Product en Consument*, Utrecht: Lemma.
- Ratneshwar, S. and, Allan D. Shocker (1988), "The Application of Prototypes and Categorization Theory in Marketing: Some Problems and Alternative Perspectives", *Advances in Consumer Research*, Vol. 15, 280-285.
- Rogers, Everett M. (1995), *Diffusion of Innovations* (4th edition), Free Press.
- Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem (1976), "Basic Objects in Natural Categories", *Cognitive Psychology*, Vol. 8, 382-439.
- Smith, Edward E., and Douglas L. Medin (1981), *Categories and Concepts*, Cambridge, MA: Harvard University Press.
- Stayman, Douglas M., Dana L. Alden, and Karen H. Smith (1992), "Some Effects of Schematic Processing on Consumer Expectations and Disconfirmation Judgments", *Journal of Consumer Research*, Vol. 19, 240-255.
- Stevens, Michael (2000), "The Essentialist Aspect of Naive Theories", *Cognition*, Vol. 74, 149-175.
- Sujan, Mita (1985), "Consumer Knowledge: Effects on Evaluation Strategies Mediating Consumer Judgments", *Journal of Consumer Research*, Vol. 12, 31-46.
- Sujan, Mita, and James R. Bettman (1989), "The Effects of Brand Positioning Strategies on Consumers' Brand and Category Perceptions: Some Insights from Schema Research", *Journal of Marketing Research*, Vol. XXVI, 454-467.

Thölke, J.M. (1998), *Product Feature Management: A Multiple Case Investigation into the Nature of Product Features and Product Feature Decision Making*, Delft: Delft University of Technology.

Tversky, Amos (1977), "Features of Similarity", *Psychological Review*, Vol. 84 (4), 327-352.

Viswanathan, Madhubalan, and Terry L. Childers (1999), "Understanding How Product Attributes Influence Product Categorization: Development and Validation of Fuzzy Set-Based Measures of Gradedness in Product Categories", *Journal of Marketing Research*, Vol. 36, 75-94.

Ward, James, and Barbara Loken (1988), "The Generality of Typicality Effects on Preference and Comparison: An Exploratory Test", *Advances in Consumer Research*, Vol. 15, 55-61.

Yi, Youjae, and Kenneth C. Gray (1996), "Revisiting Attribute Diagnosticity in the Context of Product Typicality", *Psychology and Marketing*, Vol. 13 (6), 605-632.

Ziamou, Paschalina (1998), "Consumer Learning about Discontinuous Technological Innovations: The Effect of the Degree of Newness and Information Availability", paper presented at the annual conference of the Association for Consumer Research, Montreal, October 1-4, 1998.

13. Hoe consumenten een webportaal evalueren: verschillende aspecten van on-line dienst- verlening

A.C.R. VAN RIEL, V. LILJANDER en P. JURRIËNS¹

SAMENVATTING

Het commerciële belang van diensten die aangeboden worden via internet of GSM, neemt toe. Op het vlak van de evaluatie van deze nieuwe diensten is echter nog weinig onderzoek gedaan. Er is dus dringend behoefte aan een beter begrip van de wijze waarop consumenten dit type dienstverlening evalueren, waarderen en loyaliteit ten aanzien van de dienstverlener ontwikkelen. In deze studie wordt de invloed van drie evaluatiedimensies op tevredenheid, waardeperceptie en loyaliteit vastgesteld. We verzamelden gegevens via interviews en een on-line vragenlijst. De resultaten hebben we op grond van bestaande theorieën met betrekking tot klanttevredenheid en loyaliteit geïnterpreteerd. Een aantal theoretische implicaties en suggesties voor vervolgonderzoek zijn voorzien.

1. INLEIDING

Onderzoek naar de evaluatie van dienstverlening heeft sinds het begin (Grönroos 1982; Parasuraman et al. 1985) een voortdurende ontwikkeling doorgemaakt. De hoge vlucht die het genomen heeft leidde tot menige publicatie en verschillende gespecialiseerde vaktijdschriften. Veel onderzoek werd echter uitgevoerd met betrekking tot diensten die getypeerd worden door persoonlijke interactie tussen dienstverlener en klant. Daarom is het niet zonder meer mogelijk resultaten te generaliseren naar diensten die voornamelijk gekenmerkt worden door interactie tussen consument en *technologie*, d.w.z. de vierde poot in Parasuraman's piramidemodel (Parasuraman 1997; Parasuraman 1996). Technologische ontwikkelingen in de jaren negentig hebben bedrijven in staat gesteld allerlei vernieuwingen, zoals geldautomaten en andere innovaties aan te bieden. Ook traditionele diensten worden steeds meer tevens, of zelfs uitsluitend, elektronisch aangeboden. Een recente en potentieel zeer belangrijke toevoeging aan de categorie nieuwe diensten die persoonlijke interactie vervangen en extra waarde toevoegen aan bestaande producten is het webportaal. Het is evident dat de onderzoeker op het gebied

¹ Dit hoofdstuk is gedeeltelijk gebaseerd op Van Riel et al. (2001a).

van dienstenmarketing meer aandacht zal moeten besteden aan de wijze waarop consumenten deze technologie-intensieve dienstverlening ervaren en evalueren (Parasuraman en Grewal 2000). Onderzoek naar de evaluatie van zelfbedieningstechnologie (ZBT) (Bitner et al. 2000; Dabholkar 1996; Meuter et al. 2000), waaronder verschillende vormen van elektronische dienstverlening (en on-line diensten) vallen, is nog schaars. Er is bijvoorbeeld nog weinig aandacht besteed aan de theoretische implicaties van verschillen tussen de manieren waarop klanten dienstverlening via het internet en dienstverlening in de fysieke wereld ervaren en evalueren (Grönroos et al. 2000).

1.1. Een nieuwe manier van dienstverlening: webportalen

Webportalen, te bereiken via computer (internet) of GSM (bijvoorbeeld via WAP of DoCoMo's I-mode²), zijn een volstrekt nieuwe categorie interactieve, door middel van informatietechnologie mogelijk geworden commerciële dienstenpakketten. De combinatie van hoogtechnologische ICT oplossingen en redactionele vaardigheden stelt exploitanten van webportalen in staat welbepaalde professionele segmenten te bereiken en te binden. Een webportaal bezit veel functionele voordelen, zoals bijvoorbeeld eenvoudige en snelle toegang tot onbegrensde hoeveelheden informatie en goede zoekmogelijkheden. Faciliterende en ondersteunende diensten, zoals bijvoorbeeld communicatieforums (communities) en nieuwsgroepen, kunnen bovendien extra waarde toevoegen aan het basisproduct. De exploitant kan winst maken op het basisproduct, maar ook op toegevoegde commerciële diensten, gericht op het gebruikerssegment. Deze aanvullende diensten, zoals gespecialiseerde financiële diensten, arbeidsbemiddeling (Liljander 2002), reisbureaus, of on-line bestellen van beroepsspecifieke behoeftigheden, kunnen door de verlener van de basisdienst of door derden gecommmercialiseerd worden. Webportalen zijn inmiddels een aanvulling of zelfs een alternatief geworden van traditionele gedrukte media, zoals tijdschriften en kranten, maar ook van andere media zoals radio- en televisiestations. We weten echter zeer weinig over de wijze waarop consumenten deze websites ervaren en evalueren.

Omdat er hoge kosten gemaakt moeten worden om klanten te vergaren, consumenten concurrerende diensten relatief eenvoudig op hun merites kunnen beoordelen en vergelijken, terwijl overstapkosten gering zijn, is loyaliteit van het aller grootste belang voor on-line dienstverleners (Reichheld en Scheffer 2000). Er is dus dringend onderzoek nodig met betrekking tot de rol die technologie speelt in ieder onderdeel van de ervaring van on-line dienstverlening, zoals kwaliteitsperceptie, waardeperceptie en loyaliteit (Parasuraman en Grewal 2000). Inzicht in de factoren die klanttevredenheid en loyaliteit met betrekking tot een webportaal beïnvloeden is belangrijk voor marktonderzoekers en voor managers die elektronische diensten ontwerpen en implementeren. Onze bijdrage bestaat erin, dat we een empirische studie naar klanttevredenheid met on-line dienstverlening hebben uitgevoerd. We richten ons in deze studie op een medisch webportaal dat huisartsen in ruil voor een abonnementsprijs toegang verschaft tot on-line informatie, alsmede aanvullende en ondersteunende diensten. Ons doel was het onderzoeken van het relatieve belang van bepaalde elementen van de dienstverlening voor de uiteindelijke waardeperceptie, en tevredenheid en loyaliteit ten aanzien van de dienstverlener.

We bespreken eerst huidige kennis m.b.t. evaluatie van diensten en formuleren een aan-

² Zie: <http://www.nttdocomo.com/i/index.html>

tal hypotheses die we in de empirische studie testen. Vervolgens presenteren we de opzet van die studie en de resultaten, die we interpreteren. We besluiten met enige implicaties en geven richtlijnen voor toekomstig onderzoek.

2. LITERATUURONDERZOEK

Onderzoek met betrekking tot on-line dienstverlening staat in de kinderschoenen. Er is nog geen algemeen geaccepteerd theoretisch raamwerk voor de evaluatie van deze categorie van dienstverlening voorhanden. Theorieën voor evaluatie van dienstverlening zijn in het verleden vooral ontwikkeld met het oog op het interpersoonlijke karakter van de dienstconsumptie (Meuter et al. 2000) en zijn niet of onvoldoende afgestemd op de specifieke eigenschappen van de interactie tussen klant en machine, zoals die bij on-line dienstverlening plaatsvindt. Anderzijds lijkt het onverstandig om bestaande conceptualisering van dienstenkwaliteit volledig te verwaarlozen, omdat die een solide basis kunnen bieden voor de theoretische beschrijving van on-line dienstverlening. We zien op dit ogenblik twee stromingen: één stroming propageert het voortbouwen op bestaande theorie als basis voor empirisch onderzoek (Grönroos et al. 2000; Kaynama en Black 2000; Parasuraman en Grewal 2000, Zeithaml et al. 2000), terwijl anderen een meer empirische aanpak voorstaan, hetgeen verschillende nieuwe categorieën opgeleverd heeft voor de evaluatie van ZBT (Dabholkar 1996; Meuter et al. 2000) of on-line dienstverlening (Szymanski en Hise 2000). We bestuderen eerst de proposities van beide onderzoeksrichtingen.

2.1. Concepten rond on-line dienstverlening

Door de consument ervaren kwaliteit van dienstverlening kan theoretisch weergegeven worden met behulp van twee dimensies. Berry en Parasuraman onderscheiden een proces- en een resultaatdimensie (Berry en Parasuraman 1991), terwijl Grönroos onderscheid maakt tussen functionele en technische kwaliteit (Grönroos 1990). De functionele of proceskwaliteit verwijst naar de wijze waarop de dienst verleend wordt, het 'hoe', terwijl de technische kwaliteit verwijst naar de inhoud, de voordelen die de klant ontvangt door het gebruiken van de dienst, het 'wat'. Grönroos et al. (2000) stellen voor om het traditionele dienstenbegrip, bestaande uit basisdienst, faciliterende en ondersteunende diensten, voor internetdiensten uit te breiden met een vierde factor, de gebruikersinterface. Zij stellen dat faciliterende en ondersteunende diensten moeilijk zijn te onderscheiden. Voor on-line boekhandels zijn faciliterende diensten bijvoorbeeld archieffuncties of zoekmogelijkheden en methoden om veilig on-line te kunnen betalen. Boekbesprekingen en aanbevelingen kunnen geïnterpreteerd worden als ondersteunende diensten, terwijl ontwerp, logica en functionaliteit van de site de gebruikersinterface vertegenwoordigen. Basisdienstverlening en ondersteunende diensten kunnen geïnterpreteerd worden als 'wat' de klant ontvangt, terwijl de gebruikersinterface bepaalt 'hoe' de dienst verleend en ervaren wordt.

Ondersteunende en aanvullende diensten worden gebruikt om producten te differentiëren, maar ook om de toegevoegde waarde voor de klant te verhogen (Anderson en Narus 1995; Grönroos et al. 2000). Niettemin stellen Anderson en Narus (1995) vast, dat bedrijven deze toegevoegde diensten vaak verkeerd gebruiken, door niet te onderzoeken aan welke andere diensten de klant eigenlijk behoefte heeft. Zij stellen, dat

‘klanten begrijpelijkerwijs blij zijn met gratis diensten’. ‘Maar vaak zijn aanvullende diensten niet aan de specifieke behoeften van het marktsegment waarop de basisdienst gericht is aangepast. Ze worden ongedifferentieerd en niet winstgevend aan alle gebruikers aangeboden (p. 77)’. Hoewel Anderson en Narus professionele diensten onderzocht hebben, mogen we verwachten dat hun bevindingen ook gelden voor gewone consumenten. Vaak wordt beweerd dat particuliere klanten verwachten dat alle toegevoegde diensten op het internet gratis zijn. Toch is gebleken dat lage kosten een minder belangrijke rol spelen bij het bewerkstelligen van loyaliteit dan aanvankelijk werd aangenomen (Reichheld en Schefter 2000; Hamilton 2000). Vanwege de hoge kosten, verbonden aan het onderhoud en de ontwikkeling van toegevoegde diensten, is het voor bedrijven belangrijk te bepalen of klanten ze waarderen en of ze leiden tot verhoogde loyaliteit.

We verwachten dat alle elementen van het dienstenaanbod: basisdienst, faciliterende en ondersteunende diensten en de kwaliteit van de gebruikersinterface invloed hebben op de uiteindelijke tevredenheid met een webportaal. Vanwege het ontbreken van voorgaand onderzoek op dit vlak zullen we echter geen hypothese opstellen met betrekking tot het relatieve belang van deze elementen. We vermoeden dat ze allemaal waarde toevoegen aan de basisdienst, mits ze boven de verwachtingen van de consument presteren. Die verwachtingen kunnen gebaseerd zijn op ervaringen met concurrerende on-line diensten of vergelijkbare diensten die door traditionele exploitanten aangeboden worden. Niettemin verwachten we dat de basisdienst en ondersteunende dienstverlening het meest bijdragen aan de waardeperceptie van klanten.

H1: Tevredenheid met elk element van de dienst afzonderlijk zal positief gerelateerd zijn aan algemene tevredenheid met het webportaal.

H2a: Tevredenheid met elk element van de dienst afzonderlijk zal een positief verband vertonen met perceptie van toegevoegde waarde.

H2b: Tevredenheid met de basisdienst en de ondersteunende diensten zal het sterkste effect hebben op de waardeperceptie.

Omdat toegang tot alle aangeboden diensten via de gebruikersinterface plaatsvindt, is het waarschijnlijk dat de kwaliteit hiervan de effecten van basisdienst, faciliterende diensten en ondersteunende diensten op de totale tevredenheid zal modereren. Als het bijvoorbeeld buitensporig lang duurt om bepaalde informatie te vinden, of het formaat waarin die informatie aangeboden wordt onhandig of verwarrend is, dan verwachten we dat dit een negatief effect zal hebben op de klanttevredenheid, zelfs als de informatie inhoudelijk goed is. Daarom stellen we een alternatieve hypothese op voor H1.

H3: De gebruikersinterface modereert het effect van basisdienst, faciliterende diensten en ondersteunende diensten op algemene klanttevredenheid.

2.1.1. Kwaliteitsperceptie

Elk van de vier elementen zouden we verder kunnen onderverdelen. Op het meest abstracte niveau zouden dit de proces- en resultaatdimensies kunnen zijn. Proces en resultaten van het gebruik van een zoekmachine bijvoorbeeld, van het opvragen van informatie, van het lezen van nieuwsberichten of van het bestellen van een boek. Zowel proces als resultaten van iedere interactie zouden de totale evaluatie van de on-line dienstverlening kunnen beïnvloeden. Het lijkt waarschijnlijk, dat consumenten dienstverlening als één geheel ervaren, tenzij specifieke ervaringen met één van de elementen

uitzonderlijk goed of slecht zijn. In een onderzoek naar dergelijke incidenten vonden Meuter et al. (2000) dat veel negatieve incidenten tijdens het gebruik van ZBT veroorzaakt werden door slecht functionerende processen, zoals bijvoorbeeld zoekgeraakte bestellingen of ingeslikte betaalkaarten.

Proces- en resultaatdimensies van on-line diensten kunnen we verder onderverdelen in specifiekere kwaliteitsdimensies. Bekende voorbeelden zijn de vijf SERVQUAL dimensies (Parasuraman et al. 1988): tastbaarheid, reactiesnelheid, betrouwbaarheid, vertrouwen en empathie. Parasuraman en Grewal (2000, blz. 171) stellen dat er behoefte is aan onderzoek naar hoe “definities en relatief belang van de vijf kwaliteitsdimensies veranderen als klanten met technologie te maken hebben in plaats van met personeel”. Omdat die dimensies ontwikkeld werden met het oog op evaluatie van persoonlijke confrontaties tussen klant en dienstverlener, moeten de beschrijvingen opnieuw geformuleerd worden, voordat ze zinvol gebruikt kunnen worden in een on-line dienstverleningscontext. Tastbaarheid zou bijvoorbeeld vervangen kunnen worden door de gebruikersinterface, reactiesnelheid zou betrokken kunnen worden op responsiviteit. Betrouwbaarheid zouden we in verband kunnen brengen met tijdige levering van bestellingen, accurate informatievoorziening en foutloos werkende links. Vertrouwen zou betrokken kunnen worden op de veiligheid van on-line transacties en integriteit. Vertrouwen heeft hier betrekking op het algemene vertrouwen in het bedrijf en sommige auteurs hebben geopperd dat dit één van de belangrijkste factoren is bij de bepaling van on-line tevredenheid (Petersen 2001; Urban et al. 2000; Wingfield en Rose 2001). Tenslotte zouden we empathie kunnen interpreteren als de mate waarin de communicatie op persoonlijke behoeften afgestemd is enerzijds, en de waardering voor die specifieke behoeften van de consument anderzijds. Evaluatie en relevantie van elke dimensie zullen vermoedelijk variëren voor verschillende dienstencategorieën, en wellicht zelfs per element. Consumenten zouden bijvoorbeeld wel tevreden kunnen zijn met de reactiesnelheid van de zoekmachine, maar niet met de betrouwbaarheid van de resultaten. Er kan zich ook een zekere afhankelijkheid tussen de verschillende elementen voordoen. Klanten zouden zich bijvoorbeeld tevreden kunnen stellen met tragere financiële transacties (reactiesnelheid), als dat ten goede komt aan de veiligheid (vertrouwen). Het blijft nog te zien in welke mate de vijf genoemde dimensies de wijze waarop de klant de dienstverlening ervaart volledig beschrijven, en welke mogelijke afhankelijkheden er bestaan. In deze studie zijn die dimensies niet getest.

Empirisch onderzoek naar de wijze waarop consumenten on-line dienstverlening ervaren is zeldzaam en bovendien tot op heden niet gebaseerd op bestaande theoretische conceptualisering van diensten. Er bestaan studies met betrekking tot kwaliteit van dienstverlening in specifieke zelfbedieningssituaties (Dabholkar 1996), tevredenheid en ontevredenheid met zelfbediening in het algemeen (Meuter et al. 2000), en tevredenheid met on-line diensten in het bijzonder (Szymanski en Hise 2000). Deze studies rapporteren vergelijkingen tussen ervaringen van consumenten met ZBT en hun ervaringen met persoonlijke interacties in een traditionele omgeving. Zo vonden Meuter et al. (2000) dat positieve incidenten meestal gebaseerd waren op een vergelijking met traditionele of andere ZBT, terwijl Szymanski en Hise (2000) observeerden dat gemak in vergelijking met traditionele vormen van dienstverlening een belangrijke effect had op algemene tevredenheid. Consumenten gebruiken dus zowel traditionele dienstverlening als andere ZBT als vergelijkingsmateriaal. Als voordelen van het gebruik van ZBT

worden genoemd: gemak (Meuter et al. 2000; Reichheld en Schefter 2000; Szymanski en Hise 2000), tijdbesparing en besparing van geld (Meuter et al. 2000), het vermijden van persoonlijk contact (Dabholkar 1996; Meuter et al. 2000) en het gevoel het verloop van de dienstverlening zelf in de hand te hebben (Dabholkar 1996).

Bepaalde eigenschappen van de gebruikersinterface, zoals het ontwerp van de website (Szymanski en Hise 2000), en gebruiksgemak (Dabholkar 1996; Meuter et al. 2000) beïnvloeden kwaliteitsperceptie en klanttevredenheid. Ook de indruk dat financiële transacties veilig plaats vinden beïnvloedt on-line tevredenheid (Szymanski en Hise 2000). Bovendien hebben Meuter et al. (2000) aangetoond dat klanten zich gebrekkig functionerende technologie en fouten in het design van een website wèl herinneren als negatieve incidenten, terwijl er geen verband gevonden werd tussen goed werkende technologie en positieve incidenten.

Dabholkar (1996) heeft een sterk positief verband aangetoond tussen de mate waarin consumenten plezier beleven aan het gebruik van ZBT en hun kwaliteitsperceptie. Met dit gebruiksplesier wordt de hedonistische dimensie van consumptie beschreven. Deze dimensie werd gemeten door consumenten te vragen of ze verwachtten dat de ZBT plezierig en aangenaam in het gebruik zou zijn. Overigens kan tevredenheid vanuit dit perspectief ook beschreven worden als een gematigd positieve affectieve staat (Russell 1980), maar in dienstenonderzoek wordt tevredenheid meestal beschouwd als een afhankelijke variabele die beïnvloed wordt door zowel cognitieve als affectieve reacties op de dienstverlening (Liljander en Strandvik 1997). De aanvankelijke gemoedstoestand van de klant zal bijvoorbeeld diens evaluatie van de dienstverlening beïnvloeden, terwijl het gebruik van de ZBT wellicht nieuwe affectieve reacties zal oproepen, zoals opwinding of irritatie. Buiten Dabholkar's studie is er nog geen empirisch onderzoek geweest naar affectieve reacties van klanten op ZBT. Het relatieve belang van deze factor zal waarschijnlijk afhangen van de doelgroep en het soort dienst. Hedonistische effecten zullen waarschijnlijk een belangrijkere rol spelen bij de beoordeling van diensten in verband met vrijetijdsbesteding, dan bij professionele diensten.

2.1.2. On-line loyaliteit

Loyale klanten zijn belangrijk vanwege hun bijdrage in de winstgevendheid van de dienstverlener (Anderson en Mittal 2000; Anderson en Sullivan 1993; Iacobucci et al. 1994; Storbacka et al. 1994). Ze consumeren meer dan nieuwe klanten en kunnen vaak goedkoper bediend worden. Bovendien zijn ze een gratis bron van mond tot mond reclame en betalen ze meestal de reguliere prijs (Reichheld en Sasser 1990). In een recent artikel over on-line loyaliteit observeren Reichheld en Schefter (2000) dat in tegenstelling tot gangbare veronderstellingen "niet de prijs het Internet beheerst; maar wel het vertrouwen" (blz. 107). Volgens hen zijn aanbevelingen van loyale klanten extreem belangrijk in het geval van on-line diensten. Bijna de helft van alle nieuwe klanten zou via tips van tevreden gebruikers bij een dienst terecht komen. Bovendien is gebleken dat nieuwe klanten ook advies en hulp vragen aan hun loyale voorgangers. Dit reduceert helpdeskkosten aanzienlijk.

Loyaliteit wordt weliswaar gedeeltelijk bepaald door tevredenheid³ (Reichheld en Sas-

³ In recente literatuur neemt het concept consumentenwaarde (customer value) een belangrijke plaats in. We kunnen echter aannemen dat deze waardeperceptie sterk gecorreleerd is met tevredenheid, omdat tevredenheid onder andere het resultaat is van waardeperceptie.

ser 1990; Rust et al. 1995), maar vooral het wegnemen van bronnen van ontevredenheid wordt beschouwd als cruciaal voor retentie (Anderson en Mittal 2000). Er wordt ook aangenomen dat metingen van klanttevredenheid een beter beeld geven van toekomstige financiële prestaties van diensten dan zuiver boekhoudkundige maatstaven. Anderzijds dient in ogenschouw te worden genomen, dat niet alle loyale klanten perse winstgevend zijn voor de firma. Omdat het soms jaren duurt voordat klanten winstgevend worden, is het echter in het belang van de dienstverlener om in eerste instantie te zorgen voor het behoud van zoveel mogelijk klanten, terwijl men zich daarna op de meest winstgevende klanten kan concentreren. Volgens Reichheld en Scheffer (2000), kunnen klanten alleen behouden worden door superieure kwaliteit van dienstverlening na te streven, en niet door het gebruik van technologie alleen. Daarom stellen we:

H4: Klanttevredenheid heeft een positief effect op on-line loyaliteit.

3. EMPIRISCHE STUDIE

3.1. Een webportaal

Het object van onze studie was een project van een bekende medische uitgeverij. In essentie voorziet de website⁴ zijn klanten van medische informatie door ze op de hoogte te houden van ontwikkelingen op hun vakgebied. Tijdens ons onderzoek was toegang tot de volledige inhoud gereserveerd voor betalende abonnees. Niet-betalende abonnees kregen gedurende een beperkte periode wel de gelegenheid om het volledige portaal te verkennen. Het webportaal bestond uit zeven verschillende onderdelen: nieuws, een discussieruimte, een etalage voor nieuwe publicaties, een informatiedesk, suggesties voor het organiseren van de praktijk, informatie met betrekking tot nieuwe producten en interne diensten⁵.

Hiernaast bestond de optie om de site elektronisch te doorzoeken, evenals de Amerikaanse Medline. Er werden bovendien kennistesten aangeboden met betrekking tot de verplichte opfriscursussen voor huisartsen. Dit deel was gratis. Tenslotte werd een wekelijkse nieuwsbrief verstuurd aan alle geïnteresseerden. Iedereen kon zich hierop abonneren.

Gebaseerd op het literatuuroverzicht hebben we een conceptueel raamwerk gecreëerd (zie figuur 1). Hiermee illustreren we hoe on-line dienstverlening verdeeld kan worden in vier elementen: 1) de basisdienst, 2) faciliterende diensten, 3) ondersteunende diensten en 4) de gebruikersinterface, die de klant toegang verleent tot de diensten.

De afbeelding laat zien dat consumenten zowel proces als resultaten kunnen ervaren voor alle vier de elementen en dat deze verder onderverdeeld kunnen worden in kwaliteitsdimensies. Daarnaast kunnen ze ook verschillende affectieve reacties hebben ten aanzien van elk van de afzonderlijke elementen. Terwijl de basisdienst bijvoorbeeld betrouwbaar overkomt, hoeft dat niet het geval te zijn voor de ondersteunende dienstverlening, en dat kan leiden tot frustratie en woede. Daarnaast kan de basisdienst voor bepaalde consumenten bestaan uit het kunnen vinden van informatie, terwijl dit voor

⁴ Tijdens onze studie was de prijs ongeveer EUR 100.00 per jaar, vergelijkbaar met het abonnementsgeld voor een medisch tijdschrift.

⁵ Zie Appendix 1 voor een gedetailleerde beschrijving.

Het On-Line Aanbod



Figuur 1. Conceptueel raamwerk voor on-line dienstverlening.

anderen het on-line bestellen van producten is. Voor dienstverleners is het erg belangrijk, dat *alle* onderdelen van on-line dienstverlening de klant tevredenstellen. Irritatie over onverwacht slechte ondersteunende diensten kan het vertrouwen in de basisdienst schaden.

In deze studie hebben we alleen klanttevredenheid met het oog op de drie hoofdelementen van de dienstverlening onderzocht, waarbij we faciliterende en ondersteunende diensten samengenomen hebben.

3.2. Onderzoeksopzet

De belangrijkste evaluatiecriteria van dienstverlening op het webportaal hebben we door middel van interviews met vijf betalende en voormalige gebruikers geïdentificeerd. Gebaseerd op deze gesprekken en het traditionele onderzoek naar dienstenkwaliteit hebben we een vragenlijst opgesteld.

3.3. Resultaten van interviews

Uit interviews bleek dat relevantie en betrouwbaarheid van nieuws en informatie voor huisartsen de belangrijkste factoren waren. Integriteit met betrekking tot persoonlijke gegevens vond men ook erg belangrijk. De geïnterviewden meldden, dat misbruik voor hen zelfs een reden zou zijn om hun abonnement te annuleren. Responsiviteit vond men niet zo belangrijk, omdat de huisartsen weinig belang hechtten aan direct contact of een nauwe relatie met de uitgever. Ze waren uitsluitend geïnteresseerd in contact met collega's. Aan de andere kant verwachtten ze van de uitgever dan weer wel, dat hij

moeite zou doen om op de hoogte te blijven van hun professionele interesses. Ze verwachtten dat het webportaal hen van aan hun behoeften aangepaste informatie zou voorzien. Hoofdmotief voor het gebruik van de dienst was de behoefte om op de hoogte te blijven, om voortdurend relevante en nieuwe informatie te ontvangen. Een probleem dat de meeste artsen noemden, was gebrek aan tijd om die informatie op een andere manier te verzamelen. Niemand had problemen met het inloggen en iedereen was tevreden met de internet verbinding. Ook de abonnementsprijs werd niet belangrijk gevonden, omdat die tot de fiscale aftrekposten gerekend werd. Anderen vonden de prijs onbelangrijk in zoverre het om waardevolle dienstverlening ging, terwijl een enkeling vond dat de dienstverlening gratis moest zijn.

3.4. Ontwerp van de enquête

De enquête bevatte vragen m.b.t. tevredenheid over verschillende aspecten van de site, algemene tevredenheid, voornemens om het portaal in de toekomst te gebruiken en toegevoegde waarde van verschillende aspecten van de dienstverlening. Er werden vragen toegevoegd m.b.t. potentiële verbeteringen. Alle onafhankelijke variabelen werden op 7-punts Likert schalen gemeten, lopend van 'totaal mee eens' tot 'totaal mee oneens'. De meting van algemene tevredenheid liep van 'zeer ontevreden' tot 'zeer tevreden'. Loyaliteit werd gemeten als de mate van waarschijnlijkheid dat men het portaal in de toekomst zou blijven gebruiken. Dit werd gemeten op een schaal lopend van 'zeer gering' tot 'erg groot'. Toegevoegde waarde werd gemeten als de mate waarin men het eens was met de stelling: 'Voor de huidige prijs levert het portaal goede diensten'. We vroegen de respondenten ook of zij vonden dat het portaal gratis toegankelijk moest zijn. De andere vragen behandelen we in verband met de resultaten.

De enquête werd bondig aangekondigd op de homepage van het portaal. Vanaf de aankondiging liep een link naar het webformulier dat de enquête bevatte. Iedereen die het webportaal bezocht tussen 27 november en 31 december 2000 werd zo verzocht te participeren. Tegelijkertijd werden het doel van de vragenlijst en een verzoek tot deelname aan het onderzoek in de gratis wekelijkse nieuwsbrief gepubliceerd en via e-mail verspreid. Slechts weinig abonnees vulden de vragenlijst in (response rate 15 %). Dat weinig huisartsen aan het onderzoek meededen is misschien te wijten aan de combinatie van tijdsgebrek, waar alle dokters over klaagden, en de lengte van de vragenlijst. Wellicht had het onderzoek op de homepage een meer prominente plaats kunnen krijgen en wat meer uitnodigend geformuleerd kunnen worden, maar we hadden hier weinig invloed op.

4. RESULTATEN

De meeste respondenten waren huisarts (80%) en slechts negen respondenten hadden minder dan vijf jaar werkervaring. De meesten bezochten het webportaal wekelijks, (48%) of vaker (28%) en meer dan de helft (62%) waren betalende abonnees met toegang tot alle beschikbare diensten. Het bleek dat het webportaal slechts ten dele ervaren werd als substituut voor een gedrukt tijdschrift. (3.6 op een 7-punt schaal). Redenen voor gebruik staan in tabel 1.

Tabel 1. Toegevoegde waarde van het webportaal.

Item	Aantal
1) <u>Toegevoegde waarde van het portaal naast de tijdschriften</u>	'mee eens'
* Altijd beschikbaar	39
* Opzoekmogelijkheden	36
* Archieffunctie	35
* Dagelijks nieuws	29
2) <u>Redenen voor abonnementen op het Portaal</u>	
* Op de hoogte blijven	38
* Waardevolle aanvulling	7
* Aanbevolen door collega	3

NB. Meer dan één optie kon aangekruist worden.

Onze steekproef was klein, zodat de resultaten op statistische gronden moeilijk gegeneraliseerd kunnen worden naar de volledige populatie. Anderzijds was om dezelfde redenen het arsenaal bruikbare statistische technieken uiterst beperkt. Om die reden spelen beschrijvende statistiek en correlaties een belangrijke rol. Niettemin hebben we confirmatieve regressies op een beperkt aantal variabelen uitgevoerd.

4.1. Evaluatie van het webportaal

Men verwachtte doorgaans niet dat het portaal gratis toegankelijk zou zijn (gemiddeld 2,8 op een 7-punts schaal), terwijl dit item ook niet correleerde met algemene tevredenheid of loyaliteit. Om die reden hebben we het niet verder in de analyses meegenomen. Correlatie tussen algemene tevredenheid en loyaliteit was 0,67, tussen toegevoegde waarde en tevredenheid 0,61 en tussen toegevoegde waarde en loyaliteit 0,41. In tabel 2 vermelden we correlaties tussen items uit de vragenlijst en toegevoegde waarde, algemene tevredenheid en loyaliteit. In de laatste kolom vermelden we waar het item ondergebracht kan worden: de basisdienst, de faciliterende diensten, ondersteunende diensten en/of één of meerdere kwaliteitsdimensies. We geven de items weer op volg-

Tabel 2. Correlaties tussen de perceptie van toegevoegde van het webportaal, algemene tevredenheid en loyaliteit.

Tevredenheid met elementen van de dienstverlening:	VAL	SAT	LOY	Gemiddelde	Categorie
Portaal en tijdschriften vullen elkaar goed aan	0.50	0.82	0.59	5.0	SS
Ik ontvang aanvullende informatie, in vergelijking met tijdschriften	0.36	0.75	0.51	5.2	SS
Tevredenheid met de volledigheid van de aangeboden diensten	0.39	0.69	0.41	4.9	CS
Tevredenheid over de manier waarop de diensten aangepast zijn aan mijn persoonlijke interesses	0.56	0.65	0.27	4.5	EMP
Tevredenheid over opzoekmogelijkheden	0.60	0.64	0.44	4.6	FS
Tevredenheid over het nieuwsaanbod	0.37	0.58	0.49	5.5	CS

Tevredenheid over het wetenschappelijk niveau	<u>0.60</u>	0.57	0.28	4.7	CS, REL
Tevredenheid over de toegankelijkheid	0.51	0.56	0.36	5.0	UI, RES
Tevredenheid over de wetenschappelijke inhoud	0.41	0.52	0.23	4.7	CS, REL
Tevredenheid over de logica van het portaal	0.37	0.52	0.25	5.1	UI
Tevredenheid met de voorziening van externe medische informatie	0.35	0.51	0.51	5.2	SS, EMP
Tevredenheid over advertenties m.b.t. niet-medische producten	0.47	0.51	0.31	4.6	SS
Tevredenheid over leesbaarheid van artikelen	0.46	0.49	0.41	5.4	CS, EMP
Tevredenheid met de focus van het portaal op huisartsen	0.35	0.48	0.40	5.6	CS, EMP
Tevredenheid met speciale aanbiedingen	0.34	0.48	0.44	4.1	SS
Tevredenheid over gebruiksvriendelijkheid	0.29	0.47	0.32	5.1	UI, EMP
Tevredenheid met opbouw van het portaal	0.25	0.46	0.19	4.9	UI
Tevredenheid met volgen van trends	0.33	0.43	0.25	5.3	CS, EMP
Tevredenheid met de mate waarin de informatie courant is	0.32	0.42	0.20	5.2	CS, REL
Tevredenheid met de mate van onafhankelijkheid van de industrie	0.24	0.41	0.34	5.8	REL, ASS
Tevredenheid met de wijze waarop met persoonlijk belang omgegaan wordt	0.23	0.38	0.18	4.3	EMP
Tevredenheid met de aanvoer van nieuwe artikelen	<u>0.42</u>	0.37	0.35	5.1	CS
Tevredenheid met de wijze waarop de site geïntegreerd is met andere producten	<u>0.32</u>	0.26	0.22	4.6	SS
Tevredenheid over het respect voor privacy	<u>0.35</u>	0.26	0.22	5.1	ASS
Tevredenheid met foutloze werking van links	0.17	0.19	0.26	5.1	UI, REL
Ontevredenheid met de manier waarop het portaal aangepast kan worden aan persoonlijke behoeften	0.16	0.18	<u>0.35</u>	4.1	EMP
Tevredenheid met de facturatie: één rekening voor alle diensten	0.06	0.10	<u>0.32</u>	4.8	SUPP
Tevredenheid met de mogelijkheid met andere dokters te communiceren	0.06	0.09	0.26	4.6	SS
Tevredenheid met bijstand in geval van problemen	0.26	-0.05	0.20	4.7	RES
De reputatie van de uitgever garandeert de kwaliteit van aangeboden informatie	<u>0.32</u>	0,13	0,01	5.0	REL, CS

Correlaties onder 0.28 waren niet significant. Bereik van de schalen was 1-7. VAL = toegevoegde waarde, SAT= algemene tevredenheid, LOY = intentie het portaal in de toekomst te blijven bezoeken. Categorieën: CS = basisdienst, FS= faciliterende diensten, SS= ondersteunende diensten, UI = gebruikersinterface, REL = betrouwbaarheid, RES = reactiesnelheid, ASS = vertrouwen, EMP = empathie.

orde van correlatie met algemene tevredenheid, van hoogste tot laagste waarde. Significante correlaties met toegevoegde waarde en loyaliteit zijn onderstreept wanneer ze hoger zijn dan met tevredenheid.

Uit tabel 2 volgt dat de meeste items sterker correleerden met algemene tevredenheid dan met toegevoegde waarde of loyaliteit. Dit wordt mogelijkwerwijs veroorzaakt doordat alle items met uitspraken over tevredenheid gemeten zijn, in plaats van bijvoorbeeld met kwaliteitspercepties. Hoewel dokters tijdens interviews het belang onderstreepten van een discrete behandeling van persoonlijke gegevens, had dit item geen invloed op algemene tevredenheid. Het is mogelijk, dat discretie als vanzelfsprekend wordt ervaren en dit item slechts wanneer de privacy in het gedrang komt tot ontevredenheid leidt. De wijze waarop het portaal de tijdschriften aanvult werd uitermate belangrijk gevonden voor algemene tevredenheid, maar ook voor loyaliteit. Wetenschappelijk niveau van het portaal, relevantie van de inhoud, de optie om bepaalde informatie op te zoeken en gemakkelijke toegang beïnvloedden zowel tevredenheid als toegevoegde waarde. Een andere factor die tijdens de gesprekken ter sprake kwam, namelijk de mogelijkheid om met andere huisartsen te communiceren, was niet significant gecorreleerd met algemene tevredenheid. We hebben stellig de indruk dat dit te maken heeft met de huidige nieuwsgroep-achtige organisatie van het uitwisselingsplatform, waar de discussies niet publiekelijk in real-time plaatsvinden, maar indirect via het zenden van boodschappen naar de discussieleider, die dan beslist hoe hij de boodschap publiceert. Deze optie lijkt de behoefte om direct te communiceren niet te bevredigen. We mogen ook uit tabel 3 concluderen dat directe en interactieve communicatie met vakgenoten een gewenste aanvullende dienst is.

Ontevredenheid over de wijze waarop het informatieaanbod van het portaal aansloot bij de persoonlijke interesses van de huisartsen, en tevredenheid met het opsturen van slechts één factuur voor de afname van verschillende diensten correleerden uitsluitend met loyaliteit. Vanwege de geringe omvang van de steekproef en het feit dat we ontevredenheid met slechts één item gemeten hebben, is het moeilijk om hier algemene conclusies aan te verbinden m.b.t. asymmetrie van de effecten. Toch lijkt het sterkere effect van ontevredenheid op loyaliteit aan te geven, dat asymmetrische effecten van tevredenheid en ontevredenheid op verschillende elementen van dienstverlening verder onderzocht moeten worden.

Tabel 3. Ontwikkeling van diensten op wens van de consumenten.

De kwaliteit van het portaal zou aanzienlijk verbeterd worden door de volgende dienst toe te voegen:	Gemiddelde
Reisinformatie	5.7
Mogelijkheid om contact op te nemen met vakgenoten	5.2
Een eigen homepage voor de praktijk	4.6
Mogelijkheid om software te downloaden	4.1
Informatie over goede doelen toevoegen	3.7
Financiële informatie	3.3
Mogelijkheid om medicijnen on-line te bestellen	3.1

Schaal 1-7, van 'mee oneens' tot 'mee eens'.

Een aantal potentiële ondersteunende diensten was nog niet beschikbaar, terwijl de uitgever ze wel overwoog. Bezoekers werd daarom gevraagd in hoeverre zij dachten dat

deze opties de kwaliteit zouden verbeteren. Resultaten staan in tabel 3 en tonen aan dat de belangstelling van de huisartsen voor deze opties klein was. Alleen reisinformatie en de optie om met vakgenoten te communiceren, hetzij direct of via eigen homepages, zouden substantiële effecten op de aantrekkelijkheid van het webportaal kunnen hebben.

4.2. Testen van de hypotheses

De data werden onderzocht op ontbrekende waarden, die niet gevonden werden in de gebruikte variabelen. Met behulp van de Mahalanobis D^2 methode hebben we de data op multivariate afwijkingen gecontroleerd, maar ook op dit punt hebben we geen problemen vastgesteld.

Om multivariate regressie te kunnen toepassen op onze steekproef, hebben we eerst confirmatieve factoranalyses uitgevoerd. Zo werd het aantal variabelen gereduceerd. Drie onderliggende factoren die wij identificeerden waren tevredenheid met de basisdienstverlening (CORE), tevredenheid met supplementaire dienstverlening (SUPL) en tevredenheid met de gebruikersinterface (UI). Voor de samenstelling van deze factoren, zie Appendix II.

We hebben de verdelingen van de afhankelijke variabelen getest op skewness, kurtosis en normaliteit. Geen afwijkingen werden gevonden. In een confirmatieve regressie analyse (Hair et al. 1998) hebben we de factorscores ingevoerd als onafhankelijke variabelen. Als afhankelijke variabelen dienden algemene tevredenheid met de dienstverlening (SATIS), de intentie om het webportaal opnieuw te bezoeken (LOYAL), en de ervaring van toegevoegde waarde (WAARDE). Op grond van de geringe omvang van de steekproef hebben we besloten geen interactietermen in onze regressievergelijkingen op te nemen, ofschoon we op grond van de theorie wel interactie-effecten met de tevredenheid met de gebruikersinterface kunnen verwachten. Het vereenvoudigde model kan als volgt worden weergegeven:

$$(1) \quad Y = \beta_1 + \beta_2 UI + \beta_3 CORE + \beta_4 SUPL + \varepsilon$$

Waar Y onze respectievelijke afhankelijke variabele voorstelt, β_1 de Intercept en ε een foutterm.

Omdat de nieuwe onafhankelijke variabelen betrekkelijk sterk gecorreleerd waren hebben we besloten tot een Ridge-regressie, met een K van 0.8, bij welke waarde onze coëfficiënten stabiel werden. Dit leidde tot aanvaardbare resultaten. In tabel 4 staan t -waarden en gestandaardiseerde regressiecoëfficiënten, gevonden in SPSS versie 10 (SPSS 1998). Een univariate regressie van SATIS op LOYAL werd eveneens uitgevoerd, waarvan de resultaten in tabel 5 staan.

Tabel 4. Regressiecoëfficiënten verkregen in een Multivariate Regressie.

Afhankelijke:	Tevredenheid		Loyaliteit		Toegevoegde Waarde	
Onafhankelijke:	Adj R2 .50		Adj R2 .20		Adj R2 .23	
	β	t	β	t	β	t
Intercept	5.07	44.4	5.41	27.7	4.38	20.11
Interface	.135	2.43	.031	.449	.098	1.42
Basisdienst	.203	4.01	.110	1.72	.197	3.12
Aanvullende Diensten	.290	5.82	.248	3.95	.165	2.66

Tabel 5. Regressiecoëfficiënten van Univariate Regressie: Adj R² = .35.

	Coeff. B	Std. Error	Bèta	t-waarde	Sig.
Intercept	1.367	.795		1.720	.092
SATIS	.791	.152	.600	5.197	.000

Afhankelijke Variabele: LOYAL.

Op grond van onze confirmatieve regressies aanvaardden we onze eerste hypothese: elk van de bestudeerde elementen heeft een positieve relatie met algemene tevredenheid met het webportaal. Het eerste gedeelte van onze tweede hypothese, waarin we stelden dat er een positief effect van alle onderdelen is op de ervaren toegevoegde waarde van het gebruik van het webportaal, wordt door onze data niet bevestigd. Het effect van tevredenheid over de gebruikersinterface is verwaarloosbaar. Het tweede deel van de hypothese wordt wel bevestigd, omdat basisdienst en toegevoegde diensten het sterkste effect hebben op de door de consument ervaren toegevoegde waarde. Onze derde hypothese kon niet getest worden. De vierde werd overtuigend bevestigd door de resultaten van onze univariate regressie.

5. DISCUSSIE EN SUGGESTIES VOOR VERVOLGONDERZOEK

In ons onderzoek blijkt klanttevredenheid sterk beïnvloed te worden door alle drie aspecten van on-line dienstverlening. Opmerkelijk is dat het effect van de kwaliteit en toegevoegde waarde van aanvullende diensten op tevredenheid groter is dan dat van de basisdienst. De toegevoegde waarde van aanvullende dienstverlening is blijkbaar een belangrijke driver van algemene tevredenheid en loyaliteit. Tevredenheid met de gebruikersinterface speelt ook een belangrijke rol voor wat betreft de tevredenheid zelf, maar heeft geen directe invloed op loyaliteit of ervaren waarde. De mogelijkheid bestaat echter, dat de effecten van de verschillende elementen van de dienstverlening afgezwakt worden door de ervaren kwaliteit van de gebruikersinterface. Net als in andere dienstencategorieën wordt on-line loyaliteit beïnvloed door klanttevredenheid.

Deze studie beperkte zich tot één specifieke categorie on-line dienstverlening en het empirische onderzoek had betrekking op slechts één webportaal. Verder maken snelle ontwikkelingen in de kwaliteit van on-line diensten en de kleine steekproef het moeilijk om definitieve conclusies aan dit onderzoek te verbinden. Slechts met voorzichtigheid kunnen de bevindingen gegeneraliseerd worden naar andere diensten, vergelijkbaar met de onderzochte. Er is in elk geval meer en gedetailleerder onderzoek nodig naar de determinanten van tevredenheid, loyaliteit en waardeperceptie van een hele reeks on-line diensten en andere ZBT'en.

Omdat we ons bij deze studie beperkt hebben tot het identificeren en rangschikken van onderliggende dimensies met een effect op consumentenreacties, kunnen we geen directe aanbevelingen doen met betrekking tot het verhogen van tevredenheid, toegevoegde waarde of loyaliteit. Hoewel onze resultaten wel een indicatie geven m.b.t. het belang van de ondersteunende dienstverlening als waarde-toevoegende component, is er toch meer onderzoek nodig naar de manier waarop verbeteringen bereikt kunnen

worden in de verschillende dimensies van klanttevredenheid. Omdat er bovendien belangrijke verschillen bestaan tussen evaluatiemechanismen van diensten en niet-diensten (Van Riel et al. 2001b), verwachten we ook significante verschillen tussen substantieel verschillende on-line dienstencategorieën. Een complicerende factor is, dat webportalen verschillende dienstencategorieën kunnen aanbieden aan verschillende gebruikerssegmenten. De evaluatie van dienstverlening in de verschillende dimensies en het belang hiervan voor tevredenheid en loyaliteit kan per segment verschillen.

Emoties bij het gebruik van een webportaal zijn evenmin aan bod gekomen. De rol die gevoelens spelen bij het tot stand komen van tevredenheid en loyaliteit ten aanzien van ZBT verdient zeker verder onderzocht te worden. Dabholkar (1996) stelt vast dat kwaliteitservaringen sterk gerelateerd zijn aan de mate waarin consumenten plezier beleven aan het gebruik van ZBT, terwijl Meuter et al. (2000) rapporteren dat consumenten niet direct geneigd zijn spontaan genoeg te beleven aan het gebruik van ZBT, maar dat ze wel sterke negatieve gevoelens ontwikkelen bij technische problemen. In dat opzicht zouden consumenten, wanneer de nieuwigheid van het gebruik van ZBT af is, nog slechts tevredenheid met een probleemloze dienstverlening ervaren en geen plezier meer. Negatieve emoties, voortvloeiend uit slecht functionerende dienstverlening zouden een sterker effect hebben op consumentenresponse dan positieve emoties (Liljander en Strandvik 1997), dus on-line dienstverleners doen er goed aan de kans op het ontstaan van negatieve gevoelens zo veel mogelijk te beperken, als ze hun klanten wensen te behouden. Anderzijds weten we erg weinig af van de tolerantie die consumenten hebben, of ontwikkelen en hoe negatieve gevoelens ten aanzien van ZBT hun loyaliteit beïnvloeden. Technische problemen die door de consument toegeschreven worden aan eigen computer of internetverbinding zouden bijvoorbeeld acceptabeler zijn dan technische problemen die aan de dienstverlener toegeschreven worden. Emotionele reacties van consumenten op technische problemen en de wijze waarop ze hiermee omgaan dienen verder onderzocht te worden. Als er geen werknemer beschikbaar is die men om technische bijstand kan vragen zou ontevredenheid heel snel in een exit-reactie kunnen ontaarden. Interessant in dit opzicht is dat on-line dienstverleners het blijkbaar moeilijk hebben met het volledig uitbannen van persoonlijke contacten. Meuter et al. (2000) rapporteren, dat 80 procent van de klachten over on-line dienstverlening persoonlijk bij het bedrijf werden gedeponerd, via de telefoon of door middel van een bezoek aan een kantoor.

LITERATUUR

- Anderson, Eugene W. en Claes Fornell (1999), "The Customer Satisfaction Index As A Leading Indicator," in *Handbook of Services Marketing and Management*, Dawn Iacobucci en Teresa A. Swartz, (red.) New York: Sage.
- Anderson, Eugene W. en Vital Mittal (2000), "Strengthening the Satisfaction-Profit Chain," *Journal of Service Research*, 3 (2), 107-20.
- Anderson, Eugene W. en Mary W. Sullivan (1993), "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science*, 12 (2), 125-43.
- Anderson, James C. en James A. Narus (1995), "Capturing the Value of Supplementary Services," *Harvard Business Review*, 73 (1), 75-83.
- Berry, Leonard L. en A. Parasuraman (1991), *Marketing Services: Competing through Quality*. New York: The Free Press.

- Bitner, Mary Jo, Stephen W. Brown, en Matthew L. Meuter (2000), "Technology Infusion in Service Encounters," *Journal of the Academy of Marketing Science*, 28 (1), 138-49.
- Dabholkar, Pratibha A. (1996), "Consumer Evaluations of New Technology-Based Self-Service options: An Investigation of Alternative Models," *International Journal of Research in Marketing*, 13 (1), 29-51.
- Grönroos, Christian (1990), "Relationship Approach to Marketing in Service Contexts," *Journal of Business Research*, 20 (1), 3-11.
- (1982), "Strategic Management and Marketing in the Service Sector," in Research Report. Helsinki: Swedish School of Economics and Business Administration.
- Grönroos, Christian, Fredrik Heinonen, Kristina Isoniemi, en Michael Lindholm (2000), "The Netoffer Model: A Case Example from the Virtual Marketspace," *Management Decision*, 38 (4), 243-52.
- Hair, Joseph F., Rolph E. Anderson, Ronald L. Tatham, en William C. Black (1998), *Multivariate Data Analysis* (5 ed.). Upper Saddle River, NJ: Prentice Hall, Inc.
- Hoerl, A.E. and R.W. Kennard (1970), "Ridge Regression: Biased Estimation for Non-Orthogonal Problems," *Technometrics*, 12, 55-67.
- Iacobucci, Dawn, Kent A. Grayson, en Amy L. Ostrom (1994), "The Calculus of Service Quality and Customer Satisfaction: Theoretical and Empirical Differentiation and Integration," in *Advances in Services Marketing and Management*, Teresa A. Swartz, David E. Bowen en Stephen W. Brown (red.) Vol. 3. Greenwich: JAI Press.
- Kaplan, Robert S. en David P. Norton (1996), "Using the Balanced Scorecard as a Strategic Management System," *Harvard Business Review* (January-February), 75-85.
- Kaynama, Shohreh A. en Chris I. Black (2000), "A Proposal to Assess the Service Quality of Online Travel Agencies," *Journal of Professional Services Marketing*, 21 (1), 63-68.
- Lehmann, Donald R. (1997), "A Different Game: Setting the Stage," in *A Different Game: Really New Products, Evolving Markets and Responsive Organizations*, P. Moreau (red.): Marketing Science Institute.
- Liljander, Veronica en Tore Strandvik (1997), "Emotions in Service Satisfaction," *International Journal of Service Industry Management*, 8 (2), 148-69.
- Liljander, Veronica, Allard C.R. van Riel en Minna Pura (2002), "Customer Satisfaction with E-Services: The Case of an Online Recruitment Portal," *Handbuch für Dienstleistungsmanagement 2002: Electronic Services*, Berndt Stauss, M. Bruhn (red.) Nog te verschijnen.
- Meuter, Matthew L., Amy L. Ostrom, Robert I. Roundtree, en Mary Jo Bitner (2000), "Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters," *Journal of Marketing*, 64 (July), 50-64.
- Parasuraman, A. (1997), "Augmented Marketing," *Business Today*, Annual Edition.
- (1996), "Understanding and Leveraging the Role of Customer Service in External, Interactive and Internal Marketing," in *Frontiers in Services*. Nashville TN.
- Parasuraman, A., Leonard L. Berry, en Valarie A. Zeithaml (1985), "A Conceptual Model of Service Quality and Its Implications for Future Research," *Journal of Marketing*, 49 (Fall), 41-50.
- Parasuraman, A. en Dhruv Grewal (2000), "The Impact of Technology on the Quality-Value-Loyalty Chain: A Research Agenda," *Journal of the Academy of Marketing Science*, 28 (1), 168-74.
- Parasuraman, A., Valarie A. Zeithaml, en Leonard L. Berry (1988), "SERVQUAL: A Multiple Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, 64 (1), 12-40.
- Petersen, Andrea (2001), "Private Matters. It seems that Trust Equals Revenue, even Online," in *The Wall Street Journal Europe*, February 12.
- Reichheld, Frederick F. en Earl W. Sasser (1990), "Zero Defections: Quality Comes to Services," *Harvard Business Review* (September-October), 105-11.
- Reichheld, Frederick F. en Phil Schefter (2000), "E-Loyalty: Your Secret Weapon on the Web,"

Harvard Business Review (July-August), 105-13.

Russell, James A. (1980), "A circumplex model of affect," *Journal of Personality and Social Psychology*, 39 (6), 1161-78.

Rust, Roland T., Anthony J. Zahorik, en Timothy L. Keiningham (1995), "Return on Quality (ROQ): Making Service Quality Financially Accountable," *Journal of Marketing*, 58 (April 1995), 58-70.

SPSS (1998), *SPSS Base 8.0 User's Guide*. Chicago IL: SPSS Inc.

Storbacka, Kaj (1994), "The Nature of Customer Relationship Profitability: Analysis of Relationships and Customer Bases in Retail Banking," Vol. 55. Helsinki, Finland: Publications of the Swedish School of Economics and Business Administration.

Storbacka, Kaj, Tore Strandvik en Christian Grönroos (1994), "Managing Customer Relationships for Profits: The Dynamics of Relationship Quality," *International Journal of Service Industry Management*, 5 (5), 21-38.

Szymanski, David M. en Richard T. Hise (2000), "e-Satisfaction: An Initial Examination," *Journal of Retailing*, 76 (3), 309-22.

Urban, Glen L., Fareena Sultan en William J. Qualls (2000), "Placing Trust at the Center of Your Internet Strategy," *Sloan Management Review*, 42 (1), 39-48.

Van Riel, Allard C.R., Veronica Liljander en Petra Jurriëns (2001a), "Exploring consumer evaluations of e-services: a portal site", *International Journal of Service Management*, 12 (4), 359-77.

Van Riel, Allard C.R., Jos Lemmink en Hans Ouwersloot (2001b), "Consumer Evaluations of Brand Extensions: Differences between Goods and Services," *Journal of Service Research*, 3 (3), 220-31.

Wingfield, Nick and Matthew Rose (2001), "Amazon is Split on Fees to Plug Books," in *The Wall Street Journal Europe*, February 8.

Zeithaml, Valerie A., A. Parasuraman en A. Malhotra (2000), "A Conceptual Framework for Understanding e-Service Quality: Implications for Future Research and Managerial Practice", Working Paper, Marketing Science Institute, Boston.

APPENDIX I: beschrijving van het webportaal

In *Nieuws* wordt dagelijks een aantal nieuwsberichten geplaatst. Een selectie van medisch nieuws uit kranten, tijdschriften, radio en tv wordt hier gepresenteerd, naast een archief. Dossiers worden bewaard met betrekking tot hot topics. Abonnees hebben toegang tot alle diensten. Niet-abonnees kunnen slechts titels van nieuws en archiefartikelen zien.

Op een *Discussie Forum* worden acht tot tien onderwerpen tegelijk aangeboden, plus een archief. Discussies worden via de mailinglijst aangekondigd. Dit forum kan door iedereen bezocht worden (met uitzondering van de archieven). Interactie is indirect en vindt plaats via het plaatsen van bijdrages en reacties via een discussieleider.

In *etalage* staan nieuwe medische boeken en een overzicht van alle boeken van de uitgever die het portaal exploiteert. Sommige titels kunnen on-line geraadpleegd worden en alle boeken kunnen on-line besteld worden. Zowel abonnees als niet-abonnees hebben toegang.

Gids voorziet in praktische informatie en helpt huisartsen met het wegwijs worden in de digitale medische wereld. Hier staan ongeveer 600 links naar interessante websites en een adressenbestand van allerlei medische instituten en ziekenhuizen. Tenslotte staat hier een evenementenkalender, waarop congressen, cursussen en dergelijke aangekondigd worden. Er is ook een archief met medische informatie. Het grootste deel van dit element is uitsluitend toegankelijk voor abonnees.

Praktijk is opgezet om dokters met het inrichten en organiseren van hun praktijk te helpen, terwijl tevens een poging gedaan wordt om ze bewust te maken van de vele kansen die het internet biedt. Secties in ‘*Praktijk*’ zijn: een vacatureoverzicht en een vraag en aanbod sectie voor huisartsen die bereid zijn andermans praktijk over te nemen tijdens vakanties e.d. Er wordt ook advies gegeven met betrekking tot de organisatie van financiële zaken en belastingen, hoe huisartsen ICT kunnen integreren in dagelijkse activiteiten en hoe men stress kan reduceren en meer plezier hebben in het werk.

Markt biedt nieuws m.b.t. de farmaceutische industrie en nieuwe medicijnen aan. Er staan links naar farmaceutische bedrijven en leveranciers van andere medische producten. Ook dit onderdeel is alleen toegankelijk voor abonnees.

Service bevat informatie over abonnementen, contactinformatie en advertentietarieven.

APPENDIX II: Factorsamenstelling

Tabel 1. Factorsamenstelling: Gebruikersinterface.

Gebruikersinterface (KMO: .743)
(Chi-kwadraat 110.892, Sig. .000)

Gebruiksvriendelijkheid	.877
Logica	.904
Toegankelijkheid	.760
Functionaliteit	.439
Ontwerp	.736

Tabel 2. Factorsamenstelling: Basisdienst.

Basisdienst (KMO: .805)
(Chi-kwadraat 176.085, Sig. .000)

Volledigheid	.686
Nieuws	.713
Artikelen	.705
Wetenschappelijke inhoud	.921
Wetenschappelijk niveau	.929
Leesbaarheid	.487
Beroepsmatige relevantie	.554

Tabel 3. Factorsamenstelling: Supplementaire Diensten.

Supplementaire Diensten (KMO: .805)
(Chi-kwadraat 127.332, Sig. .000)

Extra Informatie	.769
Complementaire Informatie	.809
Informatie Zoeken	.810
Externe Medische Informatie	.640
Extern niet-medisch	.728
Speciale Aanbiedingen	.819

14. Multi-niveau analyse in marktonderzoek: de case van zelfsturende service teams

A. DE JONG en K. DE RUYTER

SAMENVATTING

In het moderne marktonderzoek bestaat er steeds meer behoefte aan methoden en technieken die de onderzoeker in staat stellen om verschillende analyse niveaus met elkaar te kunnen vergelijken om zo het gedrag van mensen in groepsverband op adequate wijze te kunnen bestuderen. In dit artikel nemen we een studie naar de factoren die een rol spelen bij de totstandkoming van het service klimaat binnen zelfsturende advies teams binnen een grote Nederlandse bankorganisatie als voorbeeld van een multi-niveau onderzoek. In dit onderzoek werd een onderscheid gemaakt tussen zelfsturende teams die standaard transacties (b.v. het openen van een spaarrekening) en maatwerk services (b.v. zakelijk liquiditeitsadvies) leveren aan klanten. Daarbij is een vergelijking gemaakt tussen de percepties van medewerkers ten aanzien van de 'drivers' van het team service klimaat op het individuele niveau en het groepsniveau. Bovendien werden de medewerkers op twee verschillende tijdstippen ondervraagd. Met behulp van een hiërarchisch 'geneste' datastructuur (d.w.z., herhaalde metingen 'genest' binnen medewerkers en vervolgens medewerkers 'genest' binnen groepen) is gekeken of er sprake is van individuele effecten en/of specifieke groepseffecten van de onafhankelijke variabelen op de afhankelijke variabele team service klimaat. Uit de resultaten bleek inderdaad dat er sprake is van subtiele verschillen tussen enerzijds het meten van constructen op individueel en teamniveau en anderzijds tussen de percepties van medewerkers van standaard services en maatwerk services.

1. INLEIDING

Steeds meer organisaties gaan over tot de implementatie van zelfsturende teams om de motivatie en productiviteit van medewerkers te verhogen, kosten beter te kunnen beheersen en klanten beter van dienst te kunnen zijn (Griffin, Baldwin en Sumichrast 1994). Verschillende auteurs hebben erop gewezen dat met name in dienstverlenende organisaties zelfsturing belangrijk is, omdat de service immers tot stand komt in het contact met de klant (Larsson en Bowen 1989). Met behulp van zelfsturing in service teams kan sneller op de eisen en wensen van de klant worden ingespeeld. Om zelfsturing binnen teams op goede wijze toe te passen is het van belang dat niet alleen de kennis en vaardigheden van de teamleden elkaar aanvullen, maar ook dat functionele en sociale interacties goed verlopen (Helfert en Vith 1999). Daarom wordt de effectiviteit van een team niet alleen bepaald door eigenschappen van individuele medewerkers, maar ook door een aantal fundamentele kenmerken die betrekking hebben op het ont-

werp van het team en de processen die zich binnen het team afspelen (Spreitzer, Cohen en Ledford Jr. 1999). Uit recent onderzoek in service organisaties blijkt dat de aanwezigheid van een service klimaat sterk bepalend is voor de performance van de organisatie (Schneider, White en Paul 1998). Service klimaat heeft betrekking op de mate waarin binnen het beleid en procedures van de organisatie de service aan klanten centraal staat. Naast de implementatie van deze service orientatie op organisatie niveau, is het van belang dat er binnen zelfsturende teams eveneens een service klimaat ontstaat om het gevoerde beleid te vertalen in concrete acties die leiden tot een verhoging van de team effectiviteit (Zohar 2000).

Bij het meten van de effectiviteit van zelfsturende teams is het van belang rekening te houden met zowel het niveau van het individu als het groepsniveau. Alleen op die manier kan een onderscheid gemaakt worden tussen percepties van de individuele werknemer en gedeelde percepties op groepsniveau (Neuman en Wright 1999). De zienswijze van het individu heeft betrekking op een subjectieve inschatting, terwijl geaggregeerde gegevens van de teamleden een inter-subjectief perspectief op de werksituatie behelzen (De Jonge et al. 1999). Ostroff (1993) heeft aangetoond dat individuele scores verschillen van scores op groepsniveau. Een dergelijke bewijsvoering heeft echter weer tot een aantal andere vragen geleid. Zijn verschillen tussen niveaus gebaseerd op statistische 'artifacts', zoals meetfouten, biased schattingen, multicollineariteit of is er sprake van substantieel verschillende processen tussen beide niveaus? Om meer duidelijkheid te verschaffen hebben we een onderzoek verricht waarbij we steeds dezelfde predictor variabelen (d.w.z. factoren die van invloed zijn op het service klimaat binnen een zelfsturend team) op twee niveaus in het onderzoeksdesign meenamen. In dit artikel besteden we eerst aandacht aan het conceptuele raamwerk dat diende als uitgangspunt van de studie en de methodologische aspecten van multi-niveau onderzoek. Vervolgens doen we verslag van een onderzoek waarbij gegevens van medewerkers van twee typen zelfsturende service teams werden verzameld op twee opeenvolgende tijdstippen en is de hiërarchische data-structuur (d.w.z., herhaalde metingen 'genest' binnen medewerkers en medewerkers 'genest' binnen groepen onderzocht met gebruik van een multi-niveau analyse (Kidwell en Mossholder 1997).

2. HET THEORETISCH RAAMWERK

Er zijn de laatste tijd nogal wat onderzoeksresultaten verschenen die aantonen dat het concept service klimaat van doorslaggevend belang is voor de effectiviteit van organisaties (Andrews en Rogelberg 2001; Schneider, Ashworth, Higgs en Carr 1996; Schneider en Bowen 1985; Schneider, Wheeler en Cox 1992). Zoals gezegd heeft service klimaat betrekking op het centraal stellen van service binnen het beleid en procedures van het bedrijf. Het lijkt daarom van belang te onderzoeken hoe een dergelijke strategische focus doordringt tot de dagelijkse praktijk van medewerkers die in zelfsturende teams services verlenen. Met andere woorden, hoe wordt strategie vertaald in actie? En, welke factoren spelen daarbij een belangrijke rol?

Uit het onderzoek naar de effectiviteit van zelfsturende teams blijkt dat een aantal zaken van belang zijn (Neuman en Wright 1999; Sundstrom, De Meuse en Futrell 1990). Spreitzer et al. (1999) stellen dat een eerste vereiste voor zelfsturende teams is dat er bij het management van het bedrijf sprake is van een zekere tolerantie van zelfsturing, hetgeen inhoudt dat het team de vrijheid en verantwoordelijkheid krijgt met

betrekking tot het invullen van de werkzaamheden. Deze succesfactor heeft dus betrekking op de voedingsbodem voor zelfsturing binnen de organisatie. Daarnaast is er herhaaldelijk op gewezen dat de flexibiliteit onder teamleden om elkaars taken te kunnen overnemen bepalend is voor effectieve zelfsturing (Sundstrom et al. 1990). In de derde plaats is het zo dat zelfs teams die zelf sturing geven aan hun functioneren nog steeds afhankelijk zijn van andere teams binnen het bedrijf omdat nu eenmaal niet alles door een team kan worden gerealiseerd. Daarom is de ondersteuning door andere teams, ook wel genoemd inter-team support, van wezenlijk belang voor het team service klimaat. Hierbij gaat het dus om de service verlening tussen (zelfsturende) teams onderling. Tenslotte, is ook aangetoond dat het van belang is dat een dergelijke bereidheid tot hulp ook binnen het team plaatsvindt om de werklast evenredig te kunnen verdelen en het zg. free-riding gedrag te voorkomen (Albanese en Van Fleet 1985; Campion, Medsker en Higgs 1993; Harkins 1987). Derhalve onderscheiden we vier factoren die bepalend zouden kunnen zijn voor het service klimaat binnen zelfsturende teams: (1) zelfsturing tolerantie, (2) flexibiliteit, (3) inter-team support en (4) intra-team support. Voordat we overgaan tot de bespreking van een aantal methodologische aspecten besteden we nog even kort aandacht aan de twee typen service teams die we onderscheiden. In de literatuur over service organisaties wordt gesteld dat verschillende typen services verschillende eisen stellen aan de medewerkers (Larsson en Bowen 1989). Zo zijn bij het leveren van meer gestandaardiseerde typen services snelheid, consistentie en prijs vooral belangrijk, terwijl bij meer op maat geleverde services aandacht voor de klant en inlevingsvermogen bepalende kwaliteitscriteria zijn (Soteriou en Zenios 1999). Om de robuustheid en generaliseerbaarheid van ons model te toetsen werden in het onderzoek zelfsturende teams betrokken die gestandaardiseerde services leveren en teams die meer gericht zijn op het leveren van maatwerk en vaak meer uitgebreide vormen van dienstverlening.

3. METHODOLOGISCHE ASPECTEN VAN MULTI-NIVEAU ANALYSE

Om het hierboven kort beschreven conceptueel raamwerk te toetsen is gebruik gemaakt van multi-niveau analyses. Vaak wordt er in marketing en organisatieonderzoek gebruik gemaakt van individuele percepties om organisatie- en groepskenmerken te meten. Vervolgens worden deze individuele scores geaggregeerd (d.w.z., het gemiddelde van alle individuele scores). In ons conceptueel model hebben de variabelen betrekking op verschijnselen die zich op groeps niveau afspelen. Derhalve lijkt het consequenter om relaties tussen deze variabelen op dat niveau ook te analyseren in plaats van op het individuele niveau. Bovendien worden groepsgegevens in de literatuur als betrouwbaarder parameters en beter indicatoren beschouwd dan een enkele individuele score (Campion et al. 1993). Aggregatie van individuele scores is gerechtvaardigd als personen binnen een groep (team) op een vergelijkbare wijze antwoorden (James 1982). Met andere woorden, als de variatie in antwoorden tussen medewerkers binnen een bepaald team laag is vergeleken met de totale variatie. Vaak wordt gesteld dat systematische individuele variatie puur gebaseerd is op groeps- of organisatie-kenmerken, waarbij individuele variantie in antwoorden binnen groepen als een bron van onnauwkeurigheid en 'ruis' is beschouwd, welke synoniem is aan de error term (James 1982). Concreet gezegd, er wordt verondersteld dat individuele scores de teamscore weerge-

ven plus een random error term. Hierdoor zou het niet relevant zijn om relaties tussen groepskenmerken op de verschillende niveaus met elkaar te vergelijken. Onderzoek heeft echter aangetoond dat variantie binnen groepen niet alleen het resultaat is van random error, maar ook systematische variantie weergeeft, d.w.z. werkelijke individuele verschillen (Ostroff 1993; Van Yperen en Snijders 2000). Individuele variantie binnen groepen kan veroorzaakt zijn door daadwerkelijke onderlinge verschillen tussen medewerkers welke het resultaat zouden kunnen zijn van de wijze waarop de vragen zijn gesteld, diversiteit in demografische achtergrond, diversiteit in de medewerker-manager relaties, lage onderlinge afhankelijkheid en minder frequente interacties tussen de medewerkers binnen het team (Klein et al. 2001; Van Yperen en Snijders 2000). Bovendien blijkt uit de praktijk dat de meeste studies kleine binnen-groeps variantie ratio's vinden voor hoger-niveau constructen. Oftewel, binnen-groeps afwijkingen zijn relatief groot in vergelijking met verschillen tussen groepen. Dit impliceert de noodzaak om individuele verschillen serieus te nemen (Ostroff 1993).

Een cruciaal aspect van de analyse die betrekking heeft op groepsverbanden (zoals teams, organisaties of virtuele gemeenschappen) betreft de hiërarchische structuur. Medewerkers zijn bijvoorbeeld 'genest' binnen teams. Multi-niveau analyse is een goede methode om dergelijke datastructuren te analyseren (Hofmann 1997; Raudenbush 1993) en is in staat om de effecten van subjectieve waarnemingen (welke leiden tot *binnen-groeps* verschillen) van de effecten van groepsfactoren (leidend tot *tussen-groeps* verschillen) te scheiden. Er zijn een aantal belangrijke argumenten om rekening te houden met dit verschil in niveaus. Ten eerste, geven de verschillende analyse niveaus substantieel verschillende aspecten weer. De individuele data corresponderen met subjectieve percepties, terwijl de geaggregeerde data corresponderen met de inter-subjectieve of objectieve beoordeling van een bepaalde situatie. Dit betekent dat individuele scores niet alleen random variëren (d.w.z. dat er sprake is van een meetfout), maar ook systematisch omdat de percepties worden beïnvloed door de specifieke achtergrond en de positie van de respondent. In geaggregeerde data analyse wordt zowel de random als de systematische variabiliteit *binnen* teams buiten beschouwing gelaten. Het verwijderen van de binnen-groepsverschillen welke het gevolg zijn van meetfouten verhoogt doorgaans de objectiviteit van de data. Echter wanneer geen rekening gehouden wordt met systematische binnen-groepsafwijkingen gaat er relevante informatie verloren (Ostroff 1993). Ten tweede, eerder onderzoek heeft overtuigend aangetoond dat predictor-prestatie relaties kunnen verschillen tussen de verschillende niveaus (Bolon 1999; Ostroff 1993). Ten derde, wanneer slechts één analyse niveau wordt gebruikt is het niet mogelijk om relaties tussen variabelen te vergelijken tussen de verschillende niveaus. Een vergelijking is cruciaal bij het bepalen of er zich vergelijkbare dan wel verschillende processen afspelen op de verschillende niveaus. Daarom is het essentieel om beide niveaus te onderzoeken bij het analyseren van team service klimaat. Ten vierde, consistentie in de predictor-prestatie relaties is belangrijk bewijs van de validiteit van de percepties van de individuele medewerkers (Ostroff 1993) en vergroot het vertrouwen in de juistheid van deze percepties ten aanzien van, bijvoorbeeld groepskenmerken, wanneer de relaties op de verschillende niveaus niet wezenlijk van elkaar verschillen. Hoewel veel onderzoekers homologie van relaties op de verschillende niveaus voorstellen (Lindsley, Brass en Thomas 1995), kan ook worden gesteld dat er verschillende processen werkzaam zijn op de verschillende analyse niveaus (Ostroff 1993). Daarom verdient de vraag of *individuele* waarnemingen een resultante zijn van groepsfactoren dan wel het gevolg van de subjectieve percepties van het individu de onderzoeks aandacht (Van Yperen en Snijders 2000).

Op grond van deze aspecten besloten we om de relatie tussen de vier eerder genoemde antecedenten en service klimaat binnen zelfsturende teams te onderzoeken. Van dit onderzoek wordt in de volgende paragraaf verslag gedaan.

4. EEN EMPIRISCH ONDERZOEK

4.1. Beschrijving van de setting van het onderzoek

Leden van zelfsturende service teams van een grote Nederlandse bank participeerden in ons onderzoek. De bank is actief in zowel het zakelijke als particuliere segment en probeert d.m.v. een goede service te concurreren op een competitieve markt. Binnen elk bankfiliaal wordt een breed pakket aan diensten aangeboden bestaande uit gestandaardiseerde en maatwerk diensten zoals advies op het gebied van investeringen, huis en hypotheek services, bedrijfsconsultancy, maar ook spaarrekeningen en geldtransacties. Binnen elke bank zijn speciale teams van medewerkers verantwoordelijk voor de standaard en maatwerk services. De invoering van zelfsturing wordt beschouwd als een belangrijk veranderingsproces. Het onderzoek diende onder andere om dit proces te evalueren.

4.2. Het meetinstrument

Alle variabelen werden gemeten met een 7-punt schaal, variërend van 'sterk mee oneens' (1) tot 'sterk mee eens' (7). Er werd grotendeels gebruik gemaakt van in eerder onderzoek gevalideerde schalen. Het concept team service klimaat (7 items; ($\alpha = .87$) is gebaseerd op items speciaal ontwikkeld voor dit onderzoek aan de hand van interviews met leden van zelfsturende service teams. Zelfsturing tolerantie (6 items; ($\alpha = .90$) is grotendeels gebaseerd op een instrument ontwikkeld door (Cook et al. 1981). Inter-team support (7 items; ($\alpha = .83$) is gemeten met behulp van een schaal ontwikkeld door (Campion et al. 1993). De flexibiliteits-schaal (4 items; ($\alpha = .70$) en intra-team support-schaal (4 items; ($\alpha = .67$) waren eveneens gebaseerd op schalen ontwikkeld door Campion et al. (1993). In tabel 1 zijn enkele voorbeeld items opgenomen.

Voorts zijn confirmatorische factor analyses uitgevoerd om de psychometrische eigenschappen van de schalen vast te stellen. Het voorgestelde factor model (GFI = 0.92; AGFI = 0.91; RMSEA = 0.054; TLI = 0.91 en CFI = 0.92) fit goed met de data en vertoont unidimensionaliteit van de schalen. Vervolgens is within-method convergent validiteit onderzocht door het bekijken van de significantie en grootte van de itemladingen. Alle items laadden significant op hun respectievelijke construct waarbij 90 procent van alle items een gestandaardiseerde lading had van tenminste 0.50. Daarnaast is de discriminant validiteit nagegaan door te toetsen of paren van constructen minder dan 1.00 met elkaar gecorreleerd zijn. Dit is getoetst aan de hand van Chi-kwadrat verschiltoetsen (1 vrijheidsgraad). Alle toetsen waren significant op 5 % significantie niveau. Naast de variabelen zoals die gespecificeerd werden in het conceptuele model zijn ook een aantal achtergrond variabelen gemeten als 'opleidingsniveau', 'lengte dienstverband', 'leeftijd', 'groeps grootte' en 'percentage van eerstelijns medewerkers' opgenomen in het model als controle variabelen. Tenslotte, is de variabele 'tijdstip 2' ingevoerd om te controleren voor het tijdseffect.

Tabel 1. Voorbeeld Items.

Items	
Construct	Voorbeeld item
Team Service Klimaat	<ul style="list-style-type: none">• Ons team werkt altijd hard om de kwaliteit van de service voor onze klanten te verbeteren.
Zelfsturing tolerantie	<ul style="list-style-type: none">• In ons team mogen de collega's geheel op hun eigen oordeel afgaan bij het oplossen van problemen.
Flexibility	<ul style="list-style-type: none">• Het is gemakkelijk voor de leden van ons team om voor elkaar in te vallen.
Inter-team support	<ul style="list-style-type: none">• Ons team krijgt voldoende ondersteuning van andere teams
Intra-team support	<ul style="list-style-type: none">• Bijna alle leden van ons team leveren een gelijkwaardige bijdrage aan het werk.

4.3. Gegevensverzameling

In dit onderzoek zijn gegevens verzameld middels schriftelijke enquêtes van individuele medewerkers van zelfsturende service groepen op T_1 en op T_2 . De enquêtes zijn per post teruggestuurd naar de onderzoekers. In totaal bestond de steekproef uit 61 teams. Van de 36 maatwerk service teams werden 568 enquêtes geretourneerd (316 op T_1 en 252 op T_2) en van de standaard service teams werden 917 enquêtes (509 op T_1 en 408 op T_2) van 25 teams ontvangen.

Aan de hand van de volgende kenmerken kan de steekproef worden gekenschetst. Van de maatwerk service teams is 54.4% van de medewerkers ouder dan 40, terwijl in de gestandaardiseerde setting de meeste mensen jonger zijn dan 40 (58.7%). Met betrekking tot opleidingsniveau en lengte dienstverband blijkt dat het merendeel van de respondenten van maatwerk service teams een HBO of WO-opleiding heeft afgerond (55.6%) en een redelijk lang dienstverband achter de rug heeft bij de bank (60.6% > 5 jaar). Daarentegen hebben de meeste respondenten van de standaard service teams een MBO opleiding gevolgd (54.7%) en heeft een groot deel relatief weinig ervaring met de organisatie (48.6% < 4 jaar). De twee samples verschillen dus op een aantal belangrijke punten van elkaar.

4.4. Analyse

Om de relatie tussen team service klimaat en de antecedenten empirisch te toetsen zijn drie-niveau modellen gespecificeerd. Niveau drie betreft het groepsgemiddelde, hetgeen is gebaseerd op de gepoolde medewerker scores op T_1 en op T_2 van een team (= groepsfactor). Niveau twee geeft de gemiddelde medewerker score weer, welke is gebaseerd op de scores op T_1 en op T_2 (= individueel gemiddelde), en niveau één betreft één enkele medewerker score op een bepaald tijdstip (= individuele score). De multi-niveau modellen zijn geanalyseerd met MLwiN (Rasbash et al. 2000). Om de generaliseerbaarheid en robuustheid van het model over de verschillende service settings te toetsen zijn twee afzonderlijke drie-niveau analyses uitgevoerd met team service klimaat als de uitkomstvariabele (Spreitzer et al. 1999).

Om inter-groeps coëfficiënten en de intra-groeps coëfficiënten van de vier antecedenten met elkaar te kunnen vergelijken zijn deze opgesplitst in een groepsgemiddelde en een

intra-groeps deviatiescore (individuele score min het groepsgemiddelde)¹. De coëfficiënt van het groepsgemiddelde betreft de inter-groeps coëfficiënt, terwijl de coëfficiënt van de intra-groep deviatie scores staat voor de intra-groeps coëfficiënt (Bryk en Raudenbush 1992; Snijders en Bosker 1999). Om beide coëfficiënten te kunnen interpreteren is het van belang na te gaan of de intra-groeps deviatie coëfficiënt en de inter-groeps coëfficiënt aan elkaar gelijk zijn en wel om de volgende redenen. Ten eerste, als beide coëfficiënten significant en aan elkaar gelijk zijn functioneert de variabele op individueel niveau omdat er geen apart hoofdeffect op groepsniveau aanwezig is. Ten tweede, als de coëfficiënt voor het groepsgemiddelde significant is, maar niet significant voor de intra-groeps deviatie, dan opereert het effect alleen op groepsniveau. Ten derde, als echter de coëfficiënt voor de intra-groeps deviatievariabele significant is en niet significant voor het groepsgemiddelde dan is het effect louter gebaseerd op vergelijking van de scores *binnen* de groep. Ten vierde, als zowel het groepsgemiddelde en de intra-groeps coëfficiënt significant zijn, maar van elkaar verschillen is er sprake van een effect op individueel niveau en een specifiek effect op groepsniveau (Van Yperen en Snijders 2000).

In Tabel 2 zijn de verdeling van de variantie over de verschillende niveaus, $r_{WG(j)}$ -coëfficiënten en ICC coëfficiënten voor zelfsturing tolerantie, inter-team support, flexibiliteit, intra-team support en team service klimaat weergegeven om na te gaan in hoeverre er sprake is van intra-groeps overeenstemming. Zowel de percentages niveau-3 variantie (variërend van 2.4 % tot 19.4%) als de ICC coëfficiënten (variërend van .03 tot .23) geven aan dat voor alle variabelen een klein tot bescheiden deel inter-groeps variantie betreft. Met name voor zelfsturing tolerantie en inter-team support is een aanzienlijk deel van de totale variantie inter-groeps variantie. Opmerkelijk is dat de inter-groeps variantie voor team service klimaat in de gestandaardiseerde setting klein is in vergelijking tot de inter-groeps variantie in de maatwerk setting. Sterke standaardisatie van de verschillende service procedures kan de oorzaak zijn van de kleinere verschillen in team service klimaat tussen de verschillende teams. Verder vertonen alle variabelen een relatief hoger inter-groeps variantie voor teams die maatwerk services leveren, hetgeen er op wijst dat percepties van verschillende medewerkers binnen *hetzelfde* team meer met elkaar overeenstemmen dan in teams die standard services leveren. Voor alle variabelen betreft het grootste gedeelte van de variantie intra-groeps variantie, welke onderverdeeld is in niveau-2 (*inter*-individuele variantie) en niveau-1 (*intra*-individuele variantie). Het relatief grote aandeel van niveau-1 variantie betekent dat de medewerker percepties aanzienlijk verschillen tussen T_1 en T_2 , hetgeen deels veroorzaakt wordt door de meetfout, maar ook veroorzaakt kan zijn door veranderingen in teamsituatie door de tijd heen. De $r_{WG(j)}$ -coëfficiënten variëren van .84 tot .95 en geven aan dat de intra-groeps overeenstemming tussen de medewerkers hoog is voor alle variabelen.

Er zijn afzonderlijke analyses uitgevoerd voor de maatwerk settings en de gestandaardiseerde settings. Allereerst zijn de controle variabelen in het model ingevoerd (stap 1). Ten tweede, zijn de intra-groups deviatie scores van de vier antecedenten aan het model toegevoegd (stap 2). Tenslotte, zijn de groepsgemiddelden van de antecedenten inge-

¹ Intra-groeps deviatie is een andere benaming voor 'group-mean centering'. In alle analyses zijn niveau-1 variabelen group mean centered en niveau-3 variabelen grand-mean centered om (1) intra-groeps variantie en inter-groeps variantie van elkaar te onderscheiden, (2) de multicollineariteit te verminderen en (3) om het schatten van de modellen te vergemakkelijken (Bryk en Raudenbush 1992).

Tabel 2. Variantie van groepsfactoren (niveau-3), individuele gemiddelden (niveau-2) en individuele scores (niveau-1), $r_{WG(J)}$ coëfficiënten en ICC coëfficiënten.

Maatwerk service teams						
Variabele	Variantie in %				Coëfficiënten	
	<i>N3</i> <i>Groeps- factoren</i>	<i>N2</i> <i>Individuele gemiddelde</i>	<i>N1</i> <i>Individuele scores</i>		$r_{WG(J)}$	ICC ^b
Zelfsturing tolerantie	10.9	46.6	42.6	(11.1) ^a	.95	.12
Inter-team Support	19.4	40.2	40.4	(17.6)	.93	.23
Flexibiliteit	7.0	33.3	59.6	(34.1)	.90	.10
Intra-team Support	5.5	43.8	50.7	(36.7)	.84	.08
Team Service klimaat	7.6	45.6	46.8	(14.1)	.95	.09

Standaard service teams						
Variabele	Variantie in %				Coëfficiënten	
	<i>N3</i> <i>Groeps- factoren</i>	<i>N2</i> <i>Individuele gemiddelde</i>	<i>N1</i> <i>Individuele scores</i>		$r_{WG(J)}$	ICC ^b
Zelfsturing tolerantie	11.2	45.8	43.0	(10.4) ^a	.94	.12
Inter-team Support	9.4	44.2	46.4	(17.6)	.91	.11
Flexibiliteit	5.2	34.1	60.6	(28.5)	.84	.07
Intra-team Support	6.0	44.1	49.9	(31.5)	.84	.09
Team Service klimaat	2.4	50.1	47.6	(12.8)	.93	.03

^a Tussen haakjes de percentages intra-groepsvariantie (=niveau-1 + niveau-2 variantie) die worden toegeschreven aan error variantie =

^b ICC-coëfficiënten zijn berekend nadat de intra-groepsvariantie (=niveau-1 + niveau-2 variantie) is gecorrigeerd voor de meetfout (meetfout = $1 - \alpha$ -coëfficiënt).

voegd (stap 3). Multi-niveau modellen worden beschouwd als contextuele modellen, welke gevoelig zijn voor multicollineariteit. Daarom zijn ordinary least squares regressie analyses uitgevoerd om aan de hand van de variantie inflation factor (VIF) te controleren voor collineariteit. De VIF's van de controle variabelen en de predictor variabelen zijn respectievelijk lager dan 3.4 en 2.2. Dus, er is geen sprake van ernstige collineariteitsproblemen (Kleinbaum, Kupper en Muller 1988).

Bezien we de resultaten in tabel 3 dan vallen een aantal zaken op. Ten eerste, tonen beide analyses een significant positief effect van zelfsturing tolerantie op team service klimaat. Zowel de deviatie score als het groepsgemiddelde hebben vergelijkbare regressie coëfficiënten, hetgeen betekent dat er géén verschil is tussen het intra-groeps en inter-groeps effect². Dit betekent dat het effect in feite opereert op het individuele niveau. Ten tweede, beide analyses geven een positief effect van de intra-groeps deviatie coëfficiënt van inter-team support op team service klimaat weer, terwijl er geen effect zicht-

² Een additionele ruwe score analyse (individuele scores werden niet gecorrigeerd voor het groepsgemiddelde) was uitgevoerd om te bepalen of er significante verschillen waren tussen de intra-groeps en inter-groeps coëfficiënten van de antecedenten (Snijders en Bosker 1999).

Tabel 3. Resultaten meer-niveau analyses voor maatwerk en standaard service teams.

	Maatwerk service teams	Standaard service teams
	Coëfficiënt (se)	Coëfficiënt (se)
<i>Stap 1 (controle variabelen)</i>		
Niveau-1 coëfficiënten ^a :		
Tijdstip 2 (γ_{100})	.049 (.065)	-.082 (.057)
Opleidingsniveau (γ_{200})	-.032 (.027)	.011 (.027)
Lengte dienstverband (γ_{300})	-.038 (.038)	-.045 (.030)
Leeftijd (γ_{400})	.106 (.068)	.087 (.057)
Niveau-3 coëfficiënten ^a :		
Opleidingsniveau (γ_{001})	-.071 (.092)	-.076 (.176)
Lengte dienstverband (γ_{002})	-.203 (.115)*	-.437 (.203)*
Leeftijd (γ_{003})	.264 (.258)	.524 (.392)
Groepsgrootte (γ_{004})	.497 (.175)**	.021 (.062)
Perc. Eerstelijns mede. (γ_{005})	.116 (.353)	-.453 (.397)
Verbetering van model fit	(χ^2 (9) = 17.07*)	(χ^2 (9) = 7.53)
<i>Stap 2 (Niveau-1 predictoren)</i>		
Zelfsturing tolerantie (γ_{500})	.220 (.040)**	.205 (.030)**
Inter-team support (γ_{600})	.128 (.036)**	.100 (.028)**
Flexibiliteit (γ_{700})	.062 (.039)	.130 (.029)**
Intra-team support (γ_{800})	.237 (.035)**	.250 (.031)**
Verbetering van model fit	(χ^2 (4) = 158.51**)	(χ^2 (4) = 296.87**)
<i>Stap 3 (Niveau-3 predictoren)</i>		
Zelfsturing tolerantie (γ_{006})	.168 (.098)*	.240 (.122)*
Inter-team support (γ_{007})	-.029 (.081)	-.186 (.121)
Flexibiliteit (γ_{008})	.334 (.148)*	.062 (.134)
Intra-team support (γ_{009})	.422 (.128)**	.544 (.151)**
Verbetering van model fit:	(χ^2 (4) = 39.15**)	(χ^2 (4) = 25.82**)
Verklaarde variantie (%)		
Individuele score(N-1)	34.0	30.9
Individuele gemiddelde(N-2)	33.6	32.5
Groepsfactor(N-3)	84.5	43.5

* $p < .05$; ** $p < .01$; ^a Ongestandaardiseerde regressie coëfficiënten; ^b Standaard errors tussen haakjes. Respondenten van maatwerk service teams $N = 568$; Respondenten van standaard service teams $N=917$.

baar is op groepsniveau. De link tussen inter-team support en team service klimaat is uitsluitend gebaseerd op sociale vergelijking *binnen* de groep. Ten derde valt op dat in de maatwerk setting de groepscoëfficiënt van flexibiliteit significant positief verbonden is met team service klimaat, terwijl erop individueel niveau geen relatie bestaat. Daarentegen, vertoont de intra-groeps coëfficiënt van flexibiliteit in de gestandaardiseerde setting wél een positief effect, terwijl er geen effect van flexibiliteit is op groepsniveau. Kennelijk is in de gestandaardiseerde setting het effect van flexibiliteit uitsluitend gebaseerd op sociale vergelijking binnen de groep, terwijl bij maatwerk services het effect van flexibiliteit juist alleen op groepsniveau aanwezig is. Ten vierde, zien we dat

beide analyses significant positieve intra-groeps deviatie- en groepseffecten van intra-team support op team service klimaat vertonen. Voor de service op maat teams zijn de intra-groeps en de inter-groeps coëfficiënten ongeveer gelijk, terwijl in de standaard service settings het groepseffect significant sterker is dan het intra-groeps deviatie effect. De relatie tussen intra-team support en team service klimaat is significant sterker op groepsniveau. (Van Yperen en Snijders 2000).

5. BESPREKING VAN DE RESULTATEN

In ons onderzoek hebben we voor verschillende zelfsturende service teams onderzocht welke factoren bepalend zijn voor de totstandkoming van het service klimaat binnen het team. Onze resultaten laten positieve relaties zien tussen tolerantie van zelfsturing, flexibiliteit, inter-team support, intra-team support en team service klimaat voor beide service settings. De centrale vraag in het onderzoek was of service klimaat wordt bepaald door groepsfactoren of individuele subjectieve waarneming. Onze resultaten laten zien dat er tot op zekere hoogte sprake is van consistentie in de effecten, terwijl tegelijkertijd subtiele verschillen bloot worden gelegd. In de eerste plaats, laten we zien dat tolerantie van zelfsturing een positief effect heeft op het service klimaat binnen het team in het geval van beide typen diensten. Kijken we naar de verschillende niveaus dan zien we dat het positieve effect op het geaggregeerde niveau hetzelfde is als bij de binnen-groeps variantie. Dit wil zeggen dat de beoordeling van service team klimaat het best kan worden verklaard vanuit de subjectieve beoordeling van de medewerker. Werknemers die van mening zijn dat er een zekere vrijheid bestaat om zelf beslissingen te nemen hebben een positiever beeld van het service klimaat binnen het zelfsturende team. In de tweede plaats, varieert de relatie tussen flexibiliteit en service klimaat tussen de verschillende typen services. Bij maatwerk dienstverlening zien we alleen een positief effect op het groepsniveau. Dit betekent dat flexibiliteit vooral op het groepsniveau van invloed is op het service klimaat binnen zelfsturende teams. In teams die maatwerk moeten leveren is flexibiliteit essentieel omdat in dit soort dienstverlening diepgaande kennis van belang is. Het lijkt belangrijk dat teams bestaan uit medewerkers die over complementaire kennis en vaardigheden beschikken zodat ze relatief eenvoudig voor elkaar in kunnen springen waardoor er een service gerichte orientatie ontstaat. Bij standaard services is het effect van flexibiliteit geheel gebaseerd op de binnen groeps variantie. Dit wil zeggen dat medewerkers die in vergelijking met hun collega's meer optimistisch zijn over de flexibiliteit binnen hun team in het algemeen positiever oordelen over het service klimaat binnen hun team. In de derde plaats zien we dat inter-team support ook op positieve wijze bijdraagt aan het service klimaat. Voor beide typen services geldt dit effect op het individuele niveau. Wederom, medewerkers met een positievere blik op de ondersteuning door andere teams binnen de bank zijn positiever over het service klimaat. Ten vierde blijkt dat de positieve relatie tussen intra-team support en service klimaat varieert afhankelijk van het niveau en het type service. In het geval van maatwerk dienstverlening geldt dat de perceptie van de individuele medewerker ten aanzien van de bereidheid van collega's om behulpzaam te zijn bepalend is voor zijn perceptie van het service klimaat. Daarbij wijken de effecten op het individuele niveau niet af van het groepsniveau. Bij standaard transacties zien we echter dat het effect op het groepsniveau significant sterker is dan het effect op het individuele niveau. Het is dus zo dat de positieve bijdrage van intra-team support gebaseerd is op gedeelde percepties van teamleden én individuele

percepties. We concluderen dat de effecten op het individuele niveau het sterkst de variabele team service klimaat verklaren. Desondanks is het zo dat groepspercepties ook van belang lijken. Bij maatwerk services zien we dat er een uniek effect is op het groepsniveau voor flexibiliteit en bij standaard services zien we een dergelijk, uniek effect voor intra-team support. Op dit niveau wordt een bepaalde variantie verklaard die niet op het individuele niveau voorkomt. Bij maatwerk services is het zo dat de perceptie van flexibiliteit door het team als geheel bepalend is voor het klimaat. Dit ligt waarschijnlijk aan het feit dat er vaak sprake is van complexe en uitgebreide service verlening waarbij de beschikbaarheid van complementaire kennis en vaardigheden binnen de groep van cruciaal belang is voor het bewerkstelligen van een klimaat dat gericht is op een goede service verlening (Stevens en Campion 1999). Bij de standaard dienstverlening gaat het niet zo zeer om de beschikbaarheid van kennis en vaardigheden. Deze vorm van dienstverlening is minder complex en verloopt zelfs vaak volgens vaste scripts. Hierbij gaat het meer om de bereidheid van collega's die binnen het team bereid zijn in te springen en behulpzaam te zijn, medewerkers die als het ware 'teamplayers' zijn (Hyatt en Ruddy 1997). Ondanks het feit dat de teams die in ons onderzoek werden meegenomen tamelijk homogeen zijn (groepsleden werken bijvoorbeeld in dezelfde afdeling) blijkt toch dat er veel variantie binnen groepen is, terwijl de variantie tussen groepen relatief klein is. Daarbij komt nog dat de verschillen tussen de groepen voor met name de variabelen tolerantie van zelfsturing en inter-team support relatief groot is. Met andere woorden, voor de variabelen die het functioneren van andere teams betreffen, zijn de verschillen tussen percepties van de verschillende teamleden binnen teams relatief klein. Dit zou veroorzaakt kunnen zijn door wat in de literatuur het 'outgroup-homogeneity' effect wordt genoemd (Quattrone en Jones 1980). Medewerkers zijn in het algemeen geneigd andere teams als meer homogeen te zien dan hun eigen team, omdat ze beter op de hoogte zijn van de specifieke verschillen die zich in hun eigen groep voordoen. Tot slot zien we dat de binnen-groeps verschillen kleiner zijn voor de teams in de maatwerk setting. De medewerkers van dit soort teams zijn het sterker met elkaar eens voor wat betreft service klimaat en de antecedenten ervan.

6. AANZET TOT VERDER ONDERZOEK

De sterkte van ieder onderzoeksproject ligt in het (h)erkennen van de tekortkomingen ervan. Een dergelijk inzicht kan bijdragen aan het onderkennen van zaken voor verder onderzoek. Om te beginnen zijn onze resultaten gebaseerd op de twee verschillende typen zelfsturende service teams die werkzaam zijn binnen een en dezelfde bank. Daarom zal nader onderzoek de generaliseerbaarheid van de resultaten verder aan moeten tonen. In ons onderzoek gingen we uit van verschillende analyse niveaus. Ook werd de respons van medewerkers geëvalueerd op twee verschillende tijdstippen. Desondanks bestaat nog steeds het gevaar van de zg. 'common method variance', hetgeen tot een vertekend beeld van de resultaten zou kunnen hebben geleid. Daarom bevelen we voor vervolgonderzoek aan om gebruik te maken van verschillende gegevensbronnen. Tot slot wordt het duidelijk wanneer we alle resultaten nog eens op een rijtje zetten dat er belangrijke nuances en subtiele verschillen te onderscheiden zijn met betrekking tot de niveaus in onze analyse. In vervolgonderzoek zal het effect van andere variabelen op team service klimaat ook moeten worden geëvalueerd om te zien of de verschillen tussen de niveaus een consistent beeld laten zien.

LITERATUUR

- Andrews T.L. en S.G. Rogelberg (2001). A New Look at Service Climate: Its Relationship with Owner Service Values in Small Businesses. *Journal of Business and Psychology*, 16 (1), 119-131.
- Albanese, R. en D.D. Van Fleet (1985). Rational behavior in groups: The free-riding tendency. *Academy of Management Review*, 10, 244-255.
- Bolon D.S. (1999). Level of Analysis Considerations in Organizational Citizenship Behavior Research: An Empirical Investigation of Individual and Work Group Effects among Hospital Employees. *Health Management Research*, 12, 92-108.
- Bryk, A.S. en S.W. Raudenbush (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications.
- Campion, M.A., G.J. Medsker en A.C. Higgs (1993). Relations between Work Group Characteristics en Effectiveness: Implications for Designing Effective Work Teams. *Personnel Psychology*, 46, 823-850.
- Cook, J.D., S.J. Hepworth, T.D. Wall, en P.B. Warr (1981). *The Experience of Work: A Compendium and Review of 249 Measures and their Use*. Academic Press Inc. (Londen) Ltd.
- De Jonge, J., G.J.P. Van Breukelen, J.A. Landeweerd, en F.J.N. Nijhuis (1999). Comparing Group and Individual Level Assessments of Job Characteristics in Testing the Job Demand-control Model: A Multi-level Approach. *Human Relations*, 52, 95-122.
- Griffin, R.K., D. Baldwin en R.T. Sumichrast (1994). Self-management Information System for the Service Industry: A Conceptual Model. *Journal of Management Information Systems*, 10, 111-133.
- Harkins, S.G. (1987). Social Loafing and Social Facilitation. *Journal of Experimental Social Psychology*, 23, 1-18.
- Helfert, G. en K. Vith (1999). Relationship Marketing Teams: Improving the Utilization of Customer Relationships Potentials Through a High Team Design Quality. *Industrial Marketing Management*, 28, 553-564.
- Hofmann, D.A. (1997). An Overview of the Logic en Rationale of Hierarchical Linear Models. *Journal of Management*, 23, 723-744.
- Hyatt, D.E. en Th.M. Ruddy (1997). An Examination of the Relationship between Work Group Characteristics and Performance: Once More into the Breech. *Personnel Psychology*, 50, 553-585.
- James, L.R. (1982). Aggregation Bias in Estimates of Perceptual Agreement. *Journal of Applied Psychology*, 67, 219-229.
- Kidwell, Jr. R.E. en K.W. Mossholder (1997). Cohesiveness en Organizational Citizenship Behavior: A Multilevel Analysis Using Work Teams en Individuals. *Journal of Management*, 23, 775-793.
- Klein K.J., A.B. Conn, D.B. Smith en J.S. Sorra (2001). Is Everyone in Agreement? An Exploration of Within-Group Agreement in Employee Perceptions of the Work Environment. *Journal of Applied Psychology*, 86(1), 3-16.
- Kleinbaum, D.G., L.L. Kupper, en K.E. Muller (1988), *Applied Regression Analysis and Other Multivariable Methods* (2nd ed.). Boston: PWS-KENT Publishing Company.
- Larsson, R. en D.E. Bowen (1989). Organization and Customer: Managing, Design and Coordination of Services. *Academy of Management Review*, 14(2), 213-333.
- Lindsley, D.H., D.J. Brass en J.B. Thomas (1995). Efficacy-performance Spirals: A Multilevel Perspective. *Academy of Management Review*, 20, 645-678.
- Neuman G.A. en J. Wright (1999). Team Effectiveness: Beyond Skills and Cognitive Ability. *Journal of Applied Psychology*, 84 (3), 376-389.
- Ostroff, C. (1993). Comparing Correlations Based on Individual-Level and Aggregated Data. *Journal of Applied Psychology*, 78 (4), 569-582.
- Quattrone G.A. en E.E. Jones (1980). The Perception of Variability within In-groups and Out-

- groups: Implications for the Law of Small Numbers. *Journal of Personality and Social Psychology*, 38 (1), 141-152.
- Rasbash, J., W. Browne, H. Goldstein, M. Yang, I. Plewis, M. Healy, G. Woodhouse, D. Draper, I. Langford en T. Lewis (2000). *A User's to MLwiN Multilevel Models Project* Institute of Education, University of London.
- Raudenbush, S.W. (1993). Hierarchical Linear Models and Experimental Design. In: Edwards, L.K. (ed.), *Applied Analysis of Variatie in Behavioral Science*, 459-496. New York: Marcel Dekker.
- Schneider, B., S. Ashworth, A. Higgs en L. Carr (1996). Design, Validity, en Use of Strategically Focused Employee Attitude Surveys, *Personnel Psychology*, 49, 695-705
- Schneider, B. en D.E. Bowen (1985). Employee and Customer Perceptions of Service in Banks: Replication and Extension. *Journal of Applied Psychology*, 70, 423-433.
- Schneider, B., J.K. Wheeler en J.F. Cox (1992). A Passion for Service: Using Content Analysis to Explicate Service Climate Themes. *Journal of Applied Psychology*, 77 (5), 705-716.
- Schneider, B., S.S. White en M.C. Paul (1998). Linking Service Climate en Customer Perceptions to Service Quality: Test of a Causal Model. *Journal of Applied Psychology*, 83, 150-163.
- Snijders, T.A.B. en R.J. Bosker (1999), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publications: London, New Delhi: Thousand Oaks.
- Soteriou, A. en S.A. Zenios (1999) Operation, Quality, and Profitability in the Provision of Banking Services. *Management Science*, 45 (9), 1221-1238.
- Spreitzer, G.M., S.G. Cohen en G.E. Ledford. jr. (1999). Developing Self-Managing Work Teams in Service Organizations. *Group & Organization Management*, 24 (3), September 1999, 340-366.
- Stevens, M.J. en M.A. Campion (1999). Staffing Work Teams: Development and Validation of a Selection Test for Teamwork Settings. *Journal of Management*, 25 (2), 207-228.
- Sundstrom, E., K.P. De Meuse en D. Futrell (1990). Work Teams: Applications and Effectiveness. *American Psychologist*, 45, 120-133.
- Van Yperen, N.W. en T.A.B. Snijders (2000). A Multi-level Analysis of the Demand-Control Model: Is Stress at Work Determined by Factors at the Group Level or the Individual Level? *Journal of Occupational Health Psychology*, 5 (1), 182-190.
- Yeatts, D.E. en C.C. Hyten (1998), *High-performing Self-managed Teams: A Comparison of Theory to Practice*. Thousen Oaks, CA: Sage.
- Zohar, D. (2000). A Group-Level Model of Safety Climate: Testing the Effect of Group Climate on Microaccidents in Manufacturing Jobs. *Journal of Applied Psychology*, 85 (4), 587-596.

15. *De beïnvloeding van percepties en evaluaties van telefonische wachttijden*

P.C. VERHOEF, G. ANTONIDES en M. VAN AALST¹

SAMENVATTING

Telefonische communicatie met klanten neemt tegenwoordig een belangrijke plaats in bij veel dienstverleners. Klanten worden bij het bellen van dienstverleners vaak in de wacht gezet. Dit kan leiden tot ontevredenheid. De dienstverlener heeft allerlei mogelijkheden om de ervaringen van de klant bij het wachten te beïnvloeden. In dit onderzoek bekijken we voornamelijk het effect van drie invullingen van de wachttijd (muziek, informatie over de wachtduur en informatie over de wachtrij) op deze ervaringen. De resultaten van een experiment onder 236 consumenten laten zien, dat informatie over de wachtduur het meest effectieve instrument is. Het effect hangt echter ook af van de lengte van de wachtduur.

1. INTRODUCTIE

“Al onze medewerkers zijn momenteel in gesprek. Een ogenblik geduld alstublieft”. Iedereen die regelmatig telefonisch contact zoekt met bedrijven en instellingen is wel eens met deze boodschap geconfronteerd. Wat volgt is een vaak onzekere periode van wachten aan de telefoon, waarin niet duidelijk is wanneer men uiteindelijk verder te woord wordt gestaan.

Veel organisaties zijn zich ervan bewust dat een telefoongesprek met een (potentiële) klant een soort ‘moment van de waarheid’ is, waarin zij kunnen bewijzen hoe belangrijk die beller nu werkelijk voor hen is. Toch komt het maar al te vaak voor dat bellers in een wachtrij moeten plaatsnemen. Nederlanders gaven massaal gehoor aan de oproep van het televisieprogramma ‘Kassa’ om hun wachtergernissen te melden. Uiteindelijk resulteerde dit in een top twintig van slecht bereikbare organisaties met telefonische wachttijden variërend van vier minuten tot bijna twee uur. Hoe kwalitatief goed de uiteindelijke service ook is, het in de wacht zetten van bellers heeft in onze gehaaste maatschappij in veel gevallen een negatieve invloed op hun tevredenheid.

Toch kunnen zich situaties voordoen waarin niet kan worden voorkomen dat bellers moeten wachten aan de telefoon. Om de hieruit resulterende ontevredenheid te verlagen

¹ Dit onderzoek is uitgevoerd in het kader van het *Service Nummer College*: Een samenwerkingsverband tussen KPN Telecom en de Erasmus Universiteit. Het artikel is gebaseerd op Antonides, Gerrit, Peter C. Verhoef en Marcel van Aalst (2002), *Consumer Perceptions and Evaluations of Telephone Waiting Times*, *Journal of Consumer Psychology*, te verschijnen

staat een aantal instrumenten ter beschikking. Zo zorgt het voeren van een gratis 0800-telefoonnummer er in ieder geval voor dat de telefoonkosten van het wachten niet voor rekening komen van de beller. Organisaties kunnen echter ook proberen om de wachtervaring zelf te beïnvloeden. Vaak wordt dat gedaan door ‘invulling’ te geven aan de wachttijd in de vorm van muziek, informatie over het aantal wachtenden of informatie over de gemiddelde wachtduur. De vraag is echter of het op deze manier invullen van telefonische wachttijden het beoogde effect heeft.

In dit artikel wordt een studie gepresenteerd naar de invloed van de hierboven beschreven drie invullingen op de percepties en evaluaties van telefonische wachttijden van bellers. Na een meer algemene inleiding over (telefonische) wachttijden en de beschrijving van het conceptuele model, zullen de hypothesen worden geformuleerd. Daarna worden de gehanteerde onderzoeksmethode en analyse toegelicht. Vervolgens worden de empirische resultaten beschreven. Het geheel zal worden afgesloten met conclusies, praktische implicaties en aanbevelingen voor verder onderzoek.

2. TELEFONISCHE WACHTTIJDEN

In de marketing- en serviceliteratuur is veel onderzoek gedaan naar wachttijden. In dit onderzoek maakt men een onderscheid tussen de objectieve wachtduur en de gepercipieerde wachtduur. De objectieve wachtduur is de tijd die de consument in werkelijkheid heeft gewacht. De subjectieve wachtduur is de door consument gepercipieerde wachtduur. Er is een sterke relatie tussen de objectieve en de subjectieve wachtduur. Echter, onderzoek laat wel zien dat de consument de objectieve wachtduur gemiddeld genomen overschat (Pruyn en Smidts, 1998).

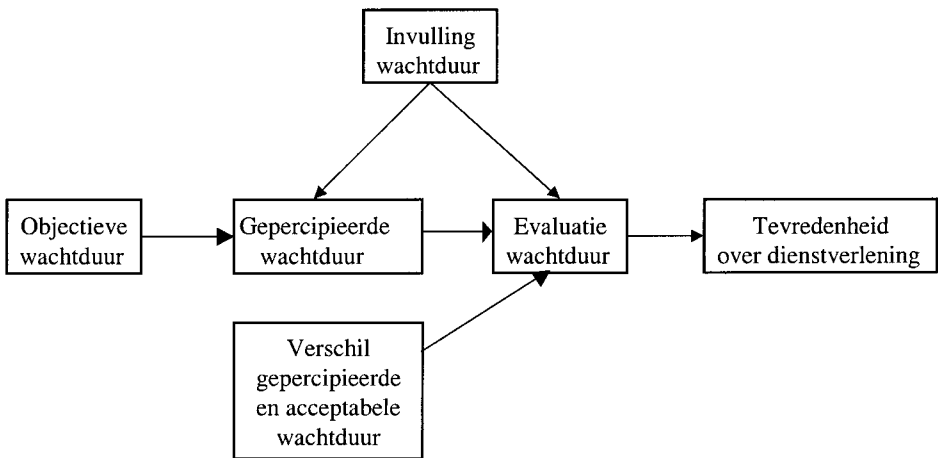
De subjectieve wachtduur beïnvloedt vervolgens de evaluatie van de wachttijd. Hierbij wordt een onderscheid gemaakt in de cognitieve evaluatie en de affectieve evaluatie. De cognitieve evaluatie betreft de vraag of consumenten het wachten te lang en/of onacceptabel vinden, terwijl de affectieve evaluatie de vraag betreft of men het wachten plezierig of vervelend vindt (Hui en Tse, 1996). Tenslotte gaat men er van uit dat de evaluatie van de wachttijd de tevredenheid over de geleverde dienst beïnvloedt. Deze tevredenheid is voor dienstverleners van wezenlijk belang, omdat men er in het algemeen vanuit gaat, dat tevreden klanten loyaler zijn (Reichheld, 1996)². Het is voor een bedrijf dus van belang om zijn wachttijden goed te managen.

Voorals in de “operations research” literatuur wordt veel aandacht besteed aan het verlagen van de objectieve wachttijden. Hier worden allerlei planningsystemen ontwikkeld om deze wachttijden tegen zo laag mogelijke kosten te minimaliseren. Het gaat bij telefonische wachttijden vooral om hoeveel call-center medewerkers op een bepaald tijdstip ingezet moeten worden. Aan dit soort planningsmethoden zullen we in dit artikel geen aandacht besteden. Er bestaan voor de dienstverlener echter ook andere methoden om de uiteindelijke tevredenheid over de telefonische dienstverlening te verhogen. Ten eerste kan de subjectieve wachtduur verlaagd worden. Ten tweede kan men trachten de evaluatie van de wachtduur te verbeteren. Bij telefonische dienstverlening zou dat kunnen door het invullen van de wachttijd (Tom et al., 1997). Dit kan men op twee manieren doen. Ten eerste kan men informatie geven over de wachtduur of het aantal wach-

² In het algemeen wordt hier wel vanuit gegaan. Onderzoek laat echter zien, dat deze relatie niet eenduidig is (zie o.a. Verhoef, Franses en Hoekstra, 2001).

tenden. Een bekend voorbeeld hiervan is het trekken van nummertjes bij een winkel. Ten tweede kan men trachten de consument tijdens het wachten te vermaken. Hiervoor worden bijvoorbeeld in wachtruimten van ziekenhuizen televisies en tijdschriften neergelegd. Bij telefonische wachttijden kan men dit doen door muziek te laten horen. In dit artikel zullen we het effect van informatie over de wachtduur, informatie over de wachtrij (het aantal wachtenden) en muziek tijdens het wachten bestuderen.

Op basis van bovenstaand overzicht leiden we dan ook het conceptuele model af zoals weergegeven in figuur 1. In dit model gaan we ervan uit, dat objectieve wachttijden de gepercipieerde wachtduur positief beïnvloeden. Deze gepercipieerde wachtduur zal vervolgens de evaluatie van het wachten negatief beïnvloeden, terwijl deze evaluatie weer een positieve invloed heeft op de tevredenheid over de dienstverlening. De invulling van de wachtduur kan de perceptie en de evaluatie van de wachtduur beïnvloeden. Dit conceptuele model geeft alleen de verwachte hoofdeffecten weer en houdt nog geen rekening met modererende effecten. In onze hypothese vorming zullen we dat wel doen. Onze hypothesen zullen vooral betrekking hebben op het effect van de invulling van wachttijden. We gaan echter ook in op het effect van de discrepantie tussen de acceptabele wachtduur en de gepercipieerde wachtduur (disconformatie) op de evaluatie van het wachten. Daarnaast zal in de analyse ook kort aandacht besteed worden aan de relatie tussen de evaluatie van de wachtduur en de tevredenheid over de dienstverlening.



Figuur 1. Het conceptuele model voor telefonische wachttijden.

3. HYPOTHESEN

Bij de hypotheseformulering zal eerst ingegaan worden op de invloed van de invulling van de wachtduur op de gepercipieerde wachtduur. Vervolgens zal het effect hiervan op de evaluatie van de wachtduur worden besproken. Tenslotte zullen nog enkele hypothesen met betrekking tot de discrepantie tussen de acceptabele wachtduur en de gepercipieerde wachtduur en de relatie tussen evaluatie van de wachtduur en tevredenheid worden behandeld.

3.1. Invulling wachtduur en gepercipieerde wachtduur

In de psychologische literatuur worden verschillende redenen aangehaald, waarom invullingen van de wachttijd de gepercipieerde wachtduur zullen beïnvloeden. In het algemeen geldt, dat complexere invullingen de gepercipieerde wachtduur verhogen, terwijl simpele invullingen de wachttijd verlagen (Hogan, 1978). Onderzoek in de services marketing literatuur laat zien dat het effect van invullingen van de wachttijd op de gepercipieerde wachtduur relatief klein is (Durrande-Moreau, 1999). Er is echter enig bewijs, dat informatie over de wachtduur leidt tot lagere inschattingen van de wachttijd (Hui and Zhou, 1996; Katz et al., 1991). Dit zou kunnen komen doordat de cognitieve inspanning verminderd wordt, omdat men zelf de wachttijd niet meer in de gaten hoeft te houden. De gegeven wachtduur zou als een soort ankerpunt gebruikt kunnen worden bij het schatten van de wachttijd, wat leidt tot een betere inschatting van de wachtduur. In de literatuur wordt geen effect van de informatie over het aantal wachtenden gevonden. Deze informatie is complexer en zal door de consument omgezet moeten worden. Dit kan een behoorlijke cognitieve inspanning zijn en zou daarom volgens de psychologische theorie zelfs kunnen leiden tot een hogere inschatting van de wachttijd. Aan de andere kant hebben consumenten wellicht ervaring opgedaan met dit soort informatie en zijn ze redelijk in staat om deze informatie naar een bepaalde tijdsinschatting te vertalen. Er wordt in eerder onderzoek geen effect van muziek op de gepercipieerde wachtduur gevonden. Wel laat het onderzoek van Pruyn en Smidts (1998) zien, dat bezoekers van een arts die televisie kijken, de wachtduur hoger inschatten. Wellicht kan dit verklaard worden doordat televisie complexe stimuli verschaft.

Op basis van bovenstaand overzicht kan dus gesteld worden, dat er voldoende theoretisch en empirisch bewijs is dat informatie over de wachtduur de gepercipieerde wachtduur negatief beïnvloedt. Wat betreft de andere twee invullingen, is zowel het empirisch als het theoretisch bewijs beperkt. We zullen hier dan ook geen hypothesen over opstellen. Onze hypothese over de invloed van de invulling van de wachtduur op de gepercipieerde wachtduur luidt dus:

H₁: Informatie over de wachtduur tijdens het wachten verlaagt de door de consument gepercipieerde wachtduur.

3.2. Invulling wachtduur en evaluatie van de wachtduur

Naar de invloed van de invulling van de wachtduur op de evaluatie ervan is tamelijk wat onderzoek gedaan. Hierbij richtte men zich vooral op het directe effect van deze invulling. We zullen dit directe effect eerst bespreken. Vervolgens zullen we echter ingaan op de veronderstelling, dat de invulling van de wachtduur ook de invloed van de lengte van de gepercipieerde wachtduur op de evaluatie van de wachtduur kan beïnvloeden.

Invulling van de wachtduur met entertainment is erop gericht de wachtduur minder onplezierig te maken. Studies laten dan ook zien dat entertainment tijdens het wachten de evaluatie van de wachtduur positief beïnvloedt (Hui et al., 1997). Wij verwachten dan ook dat muziek tijdens het wachten aan de telefoon een positief effect heeft op de evaluatie. Informatie over de wachtduur en het aantal wachtenden kan ook een positief effect hebben op de evaluatie van de wachtduur. Bellers kunnen zich door deze informatie minder gestrest voelen. Een goed voorbeeld hiervan is het wachten op een vertraagde trein. Als er niets aan de reizigers wordt verteld, zullen reizigers geïrriteerd raken, omdat men niet goed weet wanneer de trein komt. Echter, als verteld wordt dat

ze ongeveer 5 minuten moeten wachten, weet men op welke vertrektijd men ongeveer kan rekenen en stelt men zich daarop in. Een aantal onderzoeken laat dan ook positieve effecten van informatie over de wachtduur op de evaluatie van deze wachtduur zien (Hui and Tse, 1996). Volgens ons is dit effect echter complexer. Informatie over de wachtduur en wachtrij kan de beller ook meer bewust maken van de lange wachtduur. Als gevolg hiervan zal informatieverstrekking bij lange wachttijden negatief kunnen werken en zal het juist het negatieve effect van de gepercipieerde wachtduur op de evaluatie daarvan kunnen versterken. Aan de andere kant kan de informatie positief werken bij een korte wachtduur, omdat bellers juist in dat geval zich zeer bewust zijn van de korte wachttijd. Op basis van bovenstaande stellen we de volgende hypothesen op:

- H₂: Informatie over de wachtduur tijdens het wachten heeft een positieve invloed op de evaluatie van de wachtduur door de consument.*
- H₃: Informatie over het aantal wachtenden tijdens het wachten heeft een positieve invloed op de evaluatie van de wachtduur door de consument.*
- H₄: Muziek tijdens het wachten heeft een positieve invloed op de evaluatie van de wachtduur door de consument.*
- H₅: Informatie over de wachtduur tijdens het wachten versterkt het negatieve effect van de gepercipieerde wachtduur op de evaluatie van de wachtduur door de consument.*
- H₆: Informatie over het aantal wachtenden tijdens het wachten versterkt het negatieve effect van de gepercipieerde wachtduur op de evaluatie van de wachtduur door de consument.*

3.3. Discrepantie tussen gepercipieerde en acceptabele wachtduur

De service marketing literatuur laat zien, dat het verschil tussen een acceptabele prestatie en de door het bedrijf geleverde prestatie van invloed is op de evaluatie van het bedrijf (Anderson en Sullivan, 1993). Dit geldt eveneens voor het verschil tussen de gepercipieerde wachtduur en de acceptabele wachtduur (Pruyn en Smidts, 1998). Daarnaast is het belangrijk een onderscheid te maken tussen een positief verschil en een negatief verschil. Volgens de zogeheten prospect-theorie, is het absolute effect van negatieve verschillen groter dan het effect van positieve verschillen (Kahneman en Tversky, 1979). Daarnaast is het effect van deze verschillen niet-lineair, omdat het extra effect van toename van het verschil afneemt naarmate het verschil groter is. Samenvattend, het effect van een negatief verschil is convex en relatief gezien sterk, terwijl het effect van een positief verschil concaaf en relatief gezien vlak is. Om dit te modelleren kijken we dan ook naar het verschil in de logaritmes van de geaccepteerde en gepercipieerde wachtduur. Onze hypothesen zijn als volgt:

- H₇: Een negatief verschil tussen de door de consument gepercipieerde wachtduur en de acceptabele wachtduur van de consument heeft een negatief effect op de evaluatie van de wachtduur.*
- H₈: Een positief verschil tussen de door de consument gepercipieerde wachtduur en de acceptabele wachtduur van de consument heeft een positief effect op de evaluatie van de wachtduur.*
- H₉: Het effect van een negatief verschil tussen de door de consument gepercipieerde wachtduur en de acceptabele wachtduur van de consument is in absolute zin groter dan het effect van een positief verschil tussen de door de consument gepercipieerde wachtduur en de acceptabele wachtduur van de consument.*

4. ONDERZOEKSMETHODE EN ANALYSE

4.1. Methode

Om het geschetste conceptuele model en de gestelde hypothesen te testen hebben we gebruik gemaakt van een experiment onder 236 consumenten. Er is gekozen voor een experiment, omdat hiermee heel duidelijk de invloed van de bestudeerde variabelen gemeten kan worden. Een nadeel is echter dat de externe validiteit relatief klein is. Zo zullen normaal gesproken consumenten meer afleiding hebben. Het experiment vond plaats rondom winkelcentra in een aantal middelgrote en grote steden in Nederland. Op basis van een quotasteekproef met als quotavariabelen leeftijd en geslacht werden door professionele enquêteurs in winkelcentra consumenten benaderd om mee te doen aan het experiment. Een voorwaarde hierbij was wel, dat de respondent enige ervaring moest hebben met het bellen van bedrijven. Als de respondent mee wilde werken, werd hij of zij meegenomen naar een mobiele onderzoekseenheid, waarin het experiment werd uitgevoerd.

De taak van de respondent was om een nummer van de fictieve Plus Bank te bellen en om vervolgens informatie over een spaarproduct op te vragen. De experimentele variabelen in het experiment waren de wachtduur en de invulling daarvan. Door middel van een 3 x 4 factorieel design werden respondenten toegewezen aan een bepaald niveau van deze experimentele variabelen (zie tabel 1). De niveaus van de experimentele variabele wachtduur waren als volgt: 40 sec. (kort), 80 sec. (gemiddeld) en 120 sec. (lang). De niveaus waren bepaald op basis van ervaringen in een eerder experiment, waaruit bleek dat consumenten een wachttijd van rond de 60 seconden normaal vonden. De niveaus van experimentele variabele invulling waren: geen invulling van de wachtduur (stille), invulling met muziek (Titanic Theme van Celine Dion), invulling met informatie over de gemiddelde wachtduur en invulling met informatie over het aantal wachtenden. De informatie over de gemiddelde wachtduur verschilde tussen de verschillende lengtes en werd regelmatig herhaald. De gemiddelde wachtduur veranderde echter niet tijdens het wachten. De gemiddelde wachtduur zoals verteld aan de respondent was bij 40 seconden ongeveer 1 minuut, bij 80 seconden ongeveer twee minuten en bij 120 seconden ongeveer 3 minuten. De informatie over het aantal wachtenden werd wel aangepast tijdens het wachten. Naarmate een respondent langer wachtte, werd het aantal wachtenden dat gecommuniceerd werd kleiner. Het aantal wachtenden waarmee gestart werd was bij de drie wachttijden als volgt: 3 wachtenden bij 40 seconden, 6 wachtenden bij 80 seconden en 9 wachtenden bij 180 seconden. Na de wachttijd nam de telefoniste de telefoon op. Deze telefoniste beantwoordde de vragen

Tabel 1. Factorieel design experiment.

Wachtduur	Invulling wachtduur				Totaal
	Geen Invulling	Muziek	Informatie over wachtduur	Informatie over aantal wachtenden	
40 seconden	20	19	20	20	79
80 seconden	19	18	20	20	77
120 seconden	20	20	20	20	80
Totaal	59	57	60	60	236

volgens een script. Dit moest ervoor zorgen dat de telefonische beantwoording van de vragen voor elke respondent hetzelfde was. Hierdoor wordt de interne validiteit van het experiment verder verhoogd.

Na het experiment werd de respondent een vragenlijst voorgelegd. Hier moest antwoord gegeven worden op een aantal vragen over de gepercipieerde wachtduur, de acceptabele wachtduur, de evaluatie van de wachtduur, de evaluatie van de telefoniste en de tevredenheid over de telefonische dienstverlening. Deze vragen staan in tabel 2. Tenslotte werden nog kort wat vragen gesteld over het gebruik van de telefoon en demografische kenmerken, zoals leeftijd en geslacht.

De schaal voor het meten van de evaluatie van de wachtduur was betrouwbaar met een Cronbach Alpha van 0.90, terwijl de schaal ook unidimensioneel was volgens een exploratieve factor analyse. De Cronbach Alpha van de schaal om de tevredenheid over de telefoniste te meten was 0.63.

Tabel 2. Overzicht in enquête gestelde vragen.

Gepercipieerde wachtduur

Hoe lang denkt u dat u heeft moeten wachten vanaf het moment dat u hoorde dat de medewerkers in gesprek waren tot het moment dat u geholpen werd door een medewerker?seconden

Acceptabele wachtduur

Wat vindt u bij het bellen van de Plus Bank voor de informatie brochure de maximaal aanvaardbare wachttijd in seconden? seconden

Evaluatie van het wachten

Ik vond het wachten aan de telefoon (7-punts-schaal) (Bron: Hui en Tse, 1996)

- vervelend – prettig
 - saai – afwisselend
 - onbevredigend – bevredigend
 - irritant – niet irritant
 - lang – kort
 - niet acceptabel – acceptabel
-

Tevredenheid dienstverlening

Kunt u met een cijfer op een schaal van 1 tot en met 10 aangeven in welke mate u tevreden bent over het telefonisch aanvragen van de informatiebrochure bij de Plus Bank? (1 = zeer ontevreden, 10 = zeer tevreden)

Evaluatie telefoniste (1=helemaal mee oneens, 7 = helemaal mee eens)

De telefoniste gaf mij de gewenste informatie

Ik ben tevreden over de manier waarop het telefoongesprek werd afgehandeld

Ik werd netjes te woord gestaan door de telefoniste

4.2. Analyse

Het conceptuele model veronderstelt drie belangrijke variabelen: gepercipieerde wachtduur, de evaluatie van de wachtduur en de tevredenheid over de dienstverlening. Om de effecten van de verschillende variabelen op deze variabelen te testen wordt gebruik gemaakt van regressie-analyse. Hierbij moet echter wel een aantal opmerkingen worden gemaakt over de specificatie van het model. Bij het model om de gepercipieerde wacht-

duur te verklaren wordt in lijn met Antonides, Verhoef and Van Aalst (2001) een dubbel-logaritmische functie gebruikt. Dit betekent dat de afhankelijke variabele de logaritme is van de gepercipieerde wachtduur, terwijl de logaritme van de objectieve wachtduur als onafhankelijke variabele wordt gebruikt³. Om de invloed van de invulling van de wachtduur te testen worden drie dummy's opgenomen: muziek (1= invulling muziek, 0=andere invulling), informatie over wachtduur (1 = invulling informatie wachtduur, 0=andere invulling) en informatie over wachtrij (1= invulling informatie lengte wachtrij, 0 = andere invulling). Geen invulling fungeert in dit geval als de controleconditie.

In het model met de evaluatie van de wachtduur als afhankelijke variabele wordt eveneens de logaritme van de gepercipieerde wachtduur, het verschil in logaritmes van de acceptabele wachtduur en de gepercipieerde wachtduur opgenomen. Daarnaast worden ook hier weer dummy's opgenomen om het effect van het type invulling van de wachtduur te bepalen. Voor het testen van het versterkende effect van de invullingen van de wachtduur op de negatieve invloed van de gepercipieerde wachtduur worden interactie-effecten meegenomen in het model. Om het effect van de evaluatie van de wachtduur op de tevredenheid over de dienstverlening vast te stellen wordt eveneens een regressiemodel gebruikt. Hierbij wordt ook de evaluatie van de telefoniste meegenomen als antecedent van de tevredenheid over de telefonische dienstverlening.

5. EMPIRISCHE RESULTATEN

5.1. Beschrijvende statistieken

Voordat we de resultaten van de regressie-modellen beschrijven, geven we eerst kort de gemiddelde scores voor de gepercipieerde wachtduur en de evaluatie van de wachtduur weer. De gemiddelde scores staan in tabel 3. Zoals blijkt uit de tabel is er een sterk effect van de objectieve wachtduur op de gepercipieerde wachtduur. Een variantie-analyse geeft aan dat dit effect zeer significant is ($F=43$, $p<0.01$). Daarnaast is er een klein effect van de invullingen van de wachttijd op de gepercipieerde wachtduur ($F=3.00$, $p<0.05$). In de laatste drie kolommen van tabel 3 komt de evaluatie van de wachtduur aan bod. Hier zien we dat deze evaluatie daalt, als de objectieve wachtduur groter wordt ($F=20.09$; $p<0.01$). Er is er geen statistisch significant effect van de verschillende invullingen van de wachttijd ($F=1.23$, $p.0.10$).

5.2. Testresultaten

5.2.1. Gepercipieerde wachtduur

In lijn met het conceptuele model bespreken we eerst de invloed van de invulling van de wachttijden op de gepercipieerde wachtduur. De resultaten van de daarbij behorende regressie-analyse staan in tabel 4. Het regressie-model is significant en verklaart ongeveer 40% van de variantie. Zoals uit deze tabel blijkt, heeft de logaritme van de objectieve wachtduur een sterke positieve invloed op de logaritme van de gepercipieerde wachtduur. De betreffende regressie coëfficiënt is 0.84. In dit soort functies kan deze coëfficiënt opgevat worden als een elasticiteit. In dit geval betekent dit dat een stijging

³ Voor een uitgebreide argumentatie en het empirisch bewijs voor deze dubbel-logaritmische functie verwijzen wij naar Antonides et al. (2001).

Tabel 3. Gemiddelde scores voor de gepercipieerde wachtduur en de evaluatie van de wachtduur voor de verschillende condities (n=236).

	Gepercipieerde wachtduur				Evaluatie van de wachtduur			
	40 sec.	80 sec.	120 sec.	Totaal	40 sec.	80 sec.	120 sec.	Totaal
Invulling wachtduur								
Geen Invulling	83	153	207	147	4.14	3.15	2.83	3.38
Muziek	75	139	195	137	4.39	3.72	3.29	3.79
Informatie over Wachtrij	87	123	207	139	4.90	3.65	2.91	3.82
Informatie over Wachtduur	67	120	140	109	4.53	3.37	3.28	3.73
Totaal	78	133	187	133	4.49	3.47	3.08	3.68

van 100% in de objectieve wachtduur leidt tot een stijging van 84% van de subjectieve wachtduur. Met betrekking tot de invulling van de wachtduur vinden we de volgende resultaten. Muziek en informatie over de wachtrij hebben beide geen invloed op de gepercipieerde wachtduur. In lijn met hypothese 1 heeft informatie over de wachtduur een negatief effect op de gepercipieerde wachtduur. Ofwel, bellers die informatie krijgen over de wachtduur percipiëren de wachtduur als korter.

Tabel 4. Resultaten regressie-analyse invloed invulling wachtduur op logaritme gepercipieerde wachtduur (n=236).

Variabelen	Hypothese (teken)	Coëfficiënt	t-waarde
Constante		1.22	3.68**
Logaritme objectieve wachtduur		0.84	11.24**
Muziek		-0.08	0.88
Informatie over Wachtrij		-0.01	0.12
Informatie over Wachtduur	1 (-)	-0.27	2.85**

Evaluatie Model

Aangepaste $R^2 = 0.38$; F-waarde = 35**

Toelichting Tabel:

* $p < .05$. ** $p < .01$.

5.2.2. Evaluatie van de wachtduur

Om de hypothesen te testen die betrekking hebben op de evaluatie van de wachtduur schatten we twee regressie-modellen. In het eerste model zijn alleen de hoofdeffecten meegenomen. In het tweede model worden de interactie-effecten tussen de invullingen en de gepercipieerde wachtduur toegevoegd. Dit doen we, omdat we zo een beter inzicht verkrijgen. De resultaten van deze analyses staan in tabel 5. In de eerste kolom staan de onafhankelijke variabelen, terwijl in de tweede kolom de bijbehorende hypothesen staan. In de volgende twee kolommen staan de resultaten van het model met alleen de hoofdeffecten, terwijl in de laatste twee kolommen de resultaten van het model met de interactie-effecten zijn weergegeven.

Tabel 5. Resultaten regressie-analyse invloed gepercipieerde wachtduur en invulling van de wachtduur op evaluatie van de wachtduur (n=236).

Variabele	Hypothese (teken)	Alleen Hoofdeffecten		Met Interactie-Effecten	
		Coëfficiënt	t-waarde	Coëfficiënt	t-waarde
Constant		5.42	7.47**	3.95	2.95**
Logaritme Gepercipieerde Wachtduur		-0.43	2.99**	-0.12	0.44
Acceptabele Wachtduur groter dan Gepercipieerde Wachtduur (dummy)		0.31	0.96	0.36	1.09
Negatief Verschil in Logaritme acceptabele wachtduur en logaritme gepercipieerde wachtduur	7 (-)	-0.73	4.06**	-0.72	3.89**
Positief Verschil in Logaritme acceptabele wachtduur en logaritme gepercipieerde wachtduur	8 (+)	0.59	2.06*	0.50	1.70
Informatie over Wachtduur	2 (+)	0.47	2.03*	3.51	2.22*
Informatie over Wachtrij	3 (+)	0.34	1.56	2.89	1.86
Muziek	4 (+)	0.15	0.70	1.16	0.68
Informatie over Wachtduur * Logaritme Gepercipieerde Wachtduur	5 (-)			-0.72	2.17*
Informatie over Wachtrij * Logaritme Gepercipieerde Wachtduur	6 (-)			-0.54	1.67
Muziek * Logaritme Gepercipieerde Wachtduur				-0.14	0.39
<i>Model Evaluatie</i>		Aangepaste R ² = 0.44 F-waarde = 24**		Aangepaste R ² = 0.45 F-waarde = 17.81**	

Toelichting tabel:

*p < .05. ** p < .01.

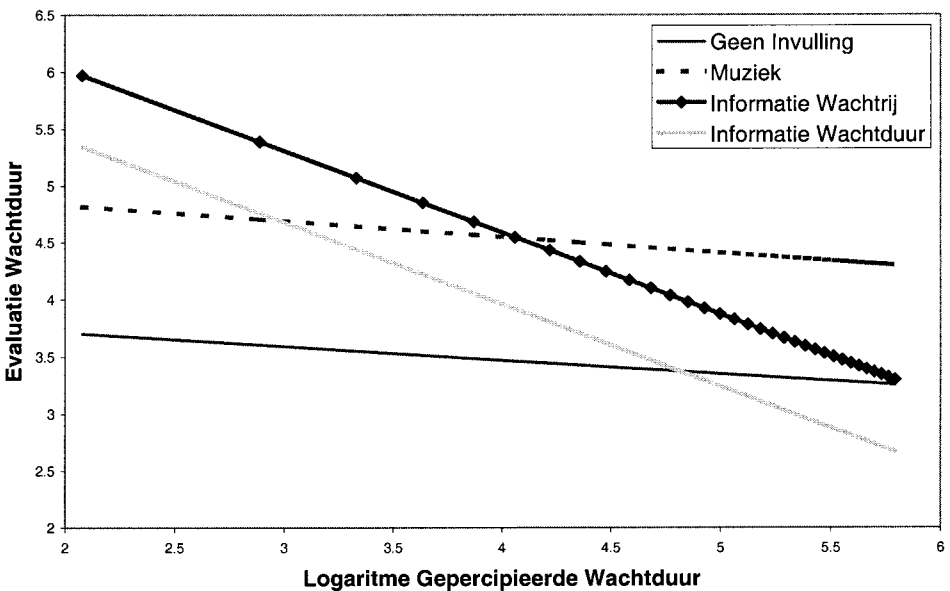
Het eerste model met alleen de hoofdeffecten is significant en de aangepaste R² is 0.44⁴. In lijn met ons conceptuele model heeft de logaritme van de gepercipieerde wachtduur een significant negatief effect (p<0.05) op de evaluatie van de wachtduur. Een negatief verschil tussen de logaritmes van de acceptabele en de gepercipieerde wachtduur heeft een significant negatieve invloed op de evaluatie (p<0.01). Hiermee wordt hypothese 7 dus ondersteund. In lijn met hypothese 8 heeft een positief verschil tussen deze twee grootheden een positieve invloed (p<0.05). Hypothese 9 veronderstelde, dat het effect van het negatieve verschil in absolute zin groter is dan het effect van

⁴ In het model is ook een dummy opgenomen voor het feit dat er een positief verschil is tussen de acceptabele en de gepercipieerde wachtduur. Dit is in lijn met Galanter (1990) die zegt dat de constante voor een positieve uitkomst hoger is dan voor een negatieve uitkomst. In dit model wordt dit niet bevestigd.

een positief verschil. Op het eerste gezicht lijkt dit het geval te zijn, omdat 0.73 inderdaad groter is dan 0.59. Als we dit echter toetsen met een Wald-test vinden we hier geen bewijs voor ($p > 0.10$) (Pindyck en Rubinfeld, 1998). Hypothese 9 kan dus niet ondersteund worden in dit experiment.

De invullingen hebben maar een beperkt effect op de evaluatie van de wachtduur. Alleen informatie over de wachtduur heeft een significant positieve invloed ($p < 0.05$). De andere invullingen hebben geen significant effect. Onze resultaten ondersteunen dan ook alleen hypothese 2, terwijl er geen bewijs wordt gevonden voor de hypothesen 3 en 4. Als in het tweede model de interactie-effecten toegevoegd worden stijgt de aangepaste R^2 tot 0.45. Zoals blijkt uit onze analyse vinden we een negatief significant interactie-effect tussen informatie over de wachtduur en de logaritme van de gepercipieerde wachtduur ($p < 0.05$). Hiermee wordt hypothese 5 ondersteund. Hypothese 6 wordt niet ondersteund, omdat de gevonden negatieve coëfficiënt niet significant is ($p > 0.05$). Tenslotte vinden we geen interactie-effect tussen muziek en de logaritme van de gepercipieerde wachtduur.

Om een beter inzicht te verkrijgen in het effect van de invullingen van de wachtduur, is op basis van de resultaten van het tweede model een grafiek gemaakt, waarin we de gepercipieerde wachtduur en de invullingen variëren. Deze grafiek is weergegeven in figuur 2. Zoals blijkt uit deze grafiek heeft informatie vooral een positieve invloed op de evaluatie van korte wachttijden. Echter, bij lange wachttijden kan dit effect zelfs negatief worden. De grens ligt hierbij rond de 90 seconden. Muziek heeft in het algemeen een positieve invloed. Zoals uit het model bleek is dit niet significant.



Figuur 2. Grafische weergave invloed invullingen en invloed gepercipieerde wachtduur op evaluatie van de wachtduur.

5.2.3. Tevredenheid dienstverlening

Tenslotte bekijken we wat de invloed van de evaluatie van de wachtduur is op de tevredenheid met de dienstverlening. Hierbij controleren we ook voor het effect van de evaluatie van de telefoniste. Het geschatte regressie-model verklaart ongeveer 40% van de variantie (zie tabel 6). Uit dit model blijkt, dat de evaluatie van de wachtduur een groot significant positief effect heeft op de tevredenheid over de dienstverlening ($p < 0.01$). Het positieve effect van de evaluatie van de telefoniste is significant ($p < 0.05$). Het is echter wel veel kleiner dan het effect van de evaluatie van de wachtduur. Of dit in werkelijkheid ook zo is, is de vraag. Door onze experimentele opzet ligt de nadruk heel erg op de wachtduur en is de aanpak van de telefoniste gestandaardiseerd. Daarnaast is het aanvragen van een brochure niet complex en kan de telefoniste dan ook weinig fout doen.

Tabel 6. Resultaten regressie-analyse invloed evaluatie wachtduur op tevredenheid telefonische dienstverlening.

Variabele	Coëfficiënt	t-waarde
Constante	0.82	0.92
Evaluatie Wachtduur	0.86	11.41**
Evaluatie Telefoniste	0.35	2.48**

Model Evaluatie

Aangepaste $R^2 = 0.40$; F-waarde = 79.73**

Toelichting Tabel

* $p < 0.05$; ** $p < 0.01$

6. SAMENVATTING, IMPLICATIES, BEPERKINGEN EN VERDER ONDERZOEK

6.1. Samenvatting en implicaties

In dit artikel is bekeken hoe de perceptie van de wachtduur bij telefonische communicatie en de evaluatie daarvan beïnvloed kan worden. Hierbij hebben we ons vooral gericht op de invulling van de wachttijd. Ons onderzoek laat zien dat bellers door het geven van informatie over de wachtduur, de wachtduur als lager percipiëren. Deze informatie over de wachtduur beïnvloedt daarnaast de evaluatie van de wachtduur positief. Echter, dit positieve effect wordt kleiner als de wachtduur langer wordt. Het effect kan dan zelfs negatief worden, omdat bellers zich meer bewust zijn van het lange wachten. Het inzetten van informatie over het wachten is bij lange wachttijden niet verstandig. Opvallend is dat we geen enkel significant effect van een muzikale invulling en informatie over de lengte van de wachtrij vinden. Kennelijk is de telefoon een te beperkt medium om met entertainment het wachten te verlichten. Al met al lijkt informatie over de wachtduur bij telefonische dienstverlening het meest effectieve instrument, omdat het zowel de gepercipieerde wachtduur als de evaluatie van de wachtduur in de juiste richting beïnvloedt. Daarnaast laat het onderzoek zien, dat het voor een dienstverlener met veel inkomend commercieel belverkeer belangrijk is om de wachtduur op een acceptabel niveau te houden. Een negatieve afwijking heeft een groot effect op de evaluatie.

6.2. Beperkingen en verder onderzoek

Dit onderzoek heeft de volgende beperkingen. Ten eerste is het een experiment. Hierdoor is de interne validiteit van de resultaten hoog, maar is de externe validiteit relatief laag. Zo worden consumenten normaal tijdens het bellen afgeleid door externe stimuli, zoals een computer, televisie of radio. Toekomstig onderzoek zou kunnen trachten data bij consumenten thuis te verzamelen. Ten tweede heeft het onderzoek alleen betrekking op het telefonisch aanvragen van informatie bij een financiële dienstverlener. Onderzoek zou zich dan ook kunnen richten op wachten in andere contexten, zoals het wachten op de trein.

Daarnaast kunnen ook nog de volgende onderzoeksmogelijkheden worden onderscheiden. Ten eerste zouden andere type invullingen, zoals commercials, of combinaties van invullingen bestudeerd kunnen worden. Ten tweede zou onderzoek zich kunnen richten op het meten van meerdere ervaringen gedurende het wachten. Hiermee kan men bekijken of deze ervaringen veranderen over de tijd en of er kritische tijden zijn waar de ervaringen zeer sterk veranderen. Tenslotte kan worden bekeken of een positief einde van het wachten leidt tot een betere waardering. Zo zou men kunnen eindigen met het geven van korting, als de klant te lang gewacht heeft.

LITERATUUR

- Anderson E.W. en M.W. Sullivan, 1993. The Antecedents and Consequences of Customer Satisfaction for Firms. *Marketing Science* 12, 125-143
- Antonides, G., P.C. Verhoef en M. van Aalst, 2002. Consumer Perception and Evaluation of Waiting Time: A Field Experiment. *Journal of Consumer Psychology* forthcoming
- Durrande-Moreau, A., 1999. Waiting for Service: Ten Years of Empirical Research. *International Journal of Service Industry Management*. 10 (2), 171-189
- Galanter, E., 1990. Utility Functions for Nonmonetary Events. *American Journal of Psychology* 103 (4), 449-470
- Hogan, H.W., 1978. A Theoretical Reconciliation of Competing Views of Time Perception. *American Journal of Psychology* 62, 423-432
- Hui, M.K. en D.K. Tse, 1996. What to Tell Consumers in Waits of Different Lengths: An Integrative Model of Service Evaluation. *Journal of Marketing* 60 (2), 81-90
- Hui M.K. en L. Zhou, 1996. How Does Waiting Duration Information Influence Customers' Reactions to Waiting for Services?. *Journal of Applied Social Psychology* 26 (19), 1702-1717
- Hui M.K., L. Dube en J.C. Chebat, 1997. The Impact of Consumers' Reaction to Waiting for Services, *Journal of Retailing* 73 (1), 87-104
- Kahneman D. en A. Tversky, 1979. Prospect Theory: An Analysis of Decision under Risk, *Econometrica* 47, 263-291
- Katz K.L., B.M. Larson en R.C. Larson, 1991. Prescription for Waiting-in-Line Blues: Entertain, Enlighten, and Engage. *Sloan Management Review*. (Winter), 44-53
- Pindyck R.S. en D.L. Rubinfeld, 1998. *Econometric Models and Economic Forecasts*. Boston: Irwin-McGraw-Hill
- Pruyn, A. en A. Smidts, 1998. Effects of Waiting on the Satisfaction with the Services: Beyond Objective Time Measures. *International Journal of Research in Marketing* 15, 321-334
- Reichheld, F.F., 1996. *The Loyalty Effect*. Cambridge MA: Harvard Business School Press
- Tom, G., M. Burns en Y. Zeng, 1997. Your Life on Hold. The Effect of Telephone Waiting Time on Customer Perception. *Journal of Direct Marketing* 11 (3), 25-31
- Verhoef P.C., P.H. Franses en J.C. Hoekstra, 2001. De Invloed van Tevredenheid op Cross-Buying: Koppeling Vragenlijst Data met de Klantendatabase. *Jaarboek Ontwikkelingen in het Marktonderzoek*, 59-72