Marktonderzoek kwartaalschrift 2-3/4

Produktonderzoeknummer

(8)

Marktonderzoek kwartaalschrift

Orgaan van de Nederlandse Vereniging van Marktonderzoekers
Secretariaat: p/a Organisatie-Bureau Wissenraet n.v.
Van Eeghenstraat 86, Amsterdam-z. Tel. 020-739551

2e jaargang, nummer 3, december 1969

Redactie J.E. Andriessens
Fr. Claasen (red. secr.)
drs. A.P. Cohen
prof. dr. ir. M. Th. G. Meulenberg
drs. H. van Stiphout
dr. A. van der Zwan

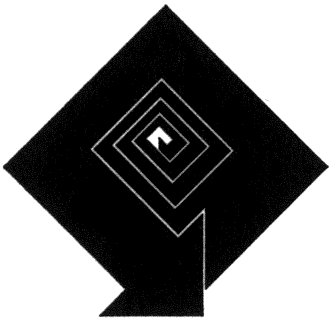
Redactie-adres John F. Kennedylaan 41, Vleuten

Uitgave Universitaire Pers Rotterdam

Administratie en advertentie-exploitatie p/a Heemraadssingel 112
Rotterdam-3
Tel. 010 - 235563

Abonnementen *f* 26,– per jaargang
studenten *f* 13,–
losse nummers *f* 7,50
incl. BTW

Auteursrecht voorbehouden



Van de redactie

VELDKAMP/MARKTONDERZOEK N.V.
Stadhouderskade 159
AMSTERDAM - 8 - Tel. 020 - 731125

De nieuwe redactie van het Marktonderzoek Kwartaalschrift stelt zich voor om steeds in een nummer van deze uitgave een bepaald onderwerp centraal te stellen waarbij naast theoretische inleidingen ook plaats is voor artikelen die gewijd zijn aan praktische toepassingen.

Als eerste poging om deze formule gestalte te geven is voor het huidige nummer het onderwerp produktonderzoek gekozen.

Het terrein van 'produktonderzoek' is bijzonder breed en alle problemen rondom dit marktonderzoekgebied komen in dit nummer zeker niet aan de orde. Bij het verzoek om bijdragen van de auteurs is er van uitgegaan dat onder produktonderzoek zou worden verstaan: onderzoek gericht op de verbetering van bestaande produkten, vergelijking van alternatieven voor een bepaald produkt, vergelijking tussen concurrerende produkten. Geen aandacht is besteed aan het testen van een volledig nieuw produkt of produktidee of aan de voorspellende waarde van de produktacceptatietest voor de positie die het produkt op de markt zal gaan innemen.

Alleen het artikel van Drs. A. W. Jongert doet verslag van een poging om de vraag naar een nieuw 'produkt' te voorspellen.

Binnen deze beperkte definitie van wat produktonderzoek is, moet de inhoud van dit nummer gelezen worden en ook dan zijn nog maar bepaalde facetten van dit type onderzoek belicht.

In het artikel van Drs. C. J. Dollé worden enkele methodologische vraagstukken besproken, terwijl de auteurs van 'Beoordeling van de kwaliteit van droogscheerapparaten' aan de hand van een case-study veel aandacht schenken aan het bijzonder belangrijke probleem van de test-opzet (test design) en de keuze van de criteria op grond waarvan de produkten beoordeeld moeten worden. Dit laatste aspect komt – op een andere wijze – ook aan de orde in het artikel van Ir. O. Grosch: 'Het vergelijkend onderzoek van gebruiksgoederen bij de Consumentenbond'.

Ten behoeve van produktontwikkeling is het veelal gewenst om de subjectieve beoordeling van de consument te relateren aan objectief meetbare eigenschappen van produkten. In het artikel van Drs. E. N. Koster wordt aan dit onderwerp aandacht geschonken.

Tenslotte kruisen de heren W. van Vliet en Prof. Dr. J. J. M. van Tulder in dit nummer de degens over het onderwerp 'Friendliness bias' naar aanleiding van de inleiding van laatstgenoemde op het ESOMAR-Congres 1969.

Ook in de volgende nummers hoopt de redactie het gekozen thema vanuit de theorie en de praktijk te kunnen laten belichten. Door de aan te snijden onderwerpen tevoren reeds aan te kondigen, rekent de redactie ook op een grotere medewerking met betrekking tot de copy voor dit tijdschrift.

Voor de komende nummers wordt gedacht aan de volgende onderwerpen:

Marktonderzoek en marktsegmentatie

Forecasting: theorie en praktijk

Onderzoek in verband met politiek: politiek gedrags- en verkiezingsonderzoek

Industrieel marktonderzoek

Naast dit deel van het tijdschrift dat de kern moet vormen, stelt de redactie zich voor de volgende rubrieken in het tijdschrift in te voeren.

Commentaar

Een rubriek die zich leent voor kortere artikelen betrekking hebbende op andere onderwerpen uit het marktonderzoek, reacties op voorgaande artikelen etc.

Uit de marktonderzoekwereld

Mededelingen van de vereniging, verslaglegging van bijeenkomsten etc.

Literatuurbespreking

Boekbesprekingen, weergave inhoud buitenlandse marktonderzoektijdschriften.

De redaktie rekent graag op uw medewerking om deze nieuwe opzet van het Marktonderzoek Kwartaalschrift te doen slagen.

A. P. Cohen

Inhoud

1 Van de redactie / drs. A. P. Cohen

3 Inhoud

4 Valstrikken in product-research / drs. C. J. Dollé

14 Het vergelijkend onderzoek van gebruiksgoederen bij de Consumentenbond / ir. O. Grosch

20 Beoordeling van de kwaliteit van droogscheerapparaten / P. Camfferman, C. P. L. Commissaris en ir. A. W. J. de Gee

33 De produkttest als brug tussen de laboratoriumwaarneming en het consumentenoordeel / drs. E. N. Koster

37 Een van de basis-problemen bij produkt-testing: de 'Friendliness-Bias'. Kritische aantekeningen bij de lezing van prof. Van Tulder tijdens het ESOMAR-Congres 1969 / W. van Vliet

45 Antwoord op de kritische aantekeningen van de heer Van Vliet / prof. dr. J. J. M. van Tulder

48 Commentaar
Hoe groot is de markt voor een nieuw produkt? / drs. A. W. Jongert

52 Uit de marktonderzoekwereld
Vaststelling per land van de totale uitgaven voor marktonderzoek. Organization on uniform lines of surveys of national expenditure on marketing research. (Draft recommendations for submission to the 113th session of the Council of the ICC)

60 NVM-Oriëntatiereis naar Amerika

61 Literatuurbespreking

67 Medewerkers

Valstrikken in product-research

Drs. C. J. Dollé

Inleiding

Er bestaan verschillende definities van het begrip 'product-research'. Ze zijn ieder voor zich aangepast aan de bedoelingen van de diverse auteurs en hebben toch ook weer veel gemeen. Dat ik aan dit aantal definities er nog één toevoeg is dan ook slechts om duidelijk te maken waarover ik het heb. In dit artikel wordt product-research opgevat als 'Het op systematische wijze nagaan in hoeverre en in welke opzichten een product wel/niet is aangepast aan de behoeften van huidige en/of mogelijke consumenten op zodanige wijze dat de verkregen kennis *desgewenst* zou kunnen leiden tot concrete maatregelen binnen het productbeleid'.

Hoewel strikt genomen de zg. 'technische' product-research, waar deze zonder consumentenoordelen werkt, toch binnen het kader van de definitie valt, blijft deze hier, vanwege het andersoortige karakter, buiten beschouwing.

Door de term 'systematisch nagaan' sluit de definitie enerzijds het gebruik van incidentele waarnemingen t.b.v. het productbeleid uit (zulke gegevens hebben slechts anekdotische waarde).

Anderzijds tracht de definitie tegen te gaan dat men de product-research zoals die tot nu toe meestal wordt bedreven, zonder meer als een volwaardige wetenschappelijke procedure beschouwt. Weliswaar bestaat het streven de product-research tot een wetenschappelijk instrument te maken, doch bij de huidige stand van zaken in de onderzoekpraktijk wordt veel product-research, hoogstens met enige systematiek, opgezet en is het slechts gebaseerd op de overwegingen van het gezonde verstand. Als zodanig kan zulk onderzoek zeer nuttige resultaten afwerpen, mits door ervaren mensen met voorzichtigheid gehanteerd.

Dergelijk onderzoek wordt echter gevaarlijk als men geavanceerde analysetechnieken zoals factoranalyse los laat op gegevens waarvan de verzameling niet aan strenge wetenschappelijke eisen voldoet. Men suggereert zodoende dat de gegevens een exactheid en een absolute geldigheid bezitten waarop ze geenszins aanspraak kunnen maken.

Daar mijn ervaring de laatste vijf jaar zich vnl. uitstrekt tot onderzoeken

m.b.t. duurzame gebruiksgoederen zullen de volgende beschouwingen vermoedelijk niet aan een zekere eenzijdigheid ontsnappen. Ik heb echter alle vertrouwen dat veel ervan toch overdraagbaar is op andere terreinen van product-research of marktonderzoek überhaupt. Ieder kan dit voor zijn eigen terrein nagaan volgens het aloude gezegde: 'Wie de schoen past, trekke hem aan'.

Om enige orde in de verwarrende veelheid van mogelijke valstrikken in product-research te scheppen, wordt uitgegaan van een indeling, die uiteraard evenals andere mogelijke indelingen betrekkelijk willekeurig is, t.w. valstrikken van:

- statistische aard
- maat-theoretische aard
- semantische aard

Uit de te behandelen voorbeelden zal blijken dat valstrikken soms onder meer dan één van de bovengenoemde rubrieken kunnen vallen, eens te meer een reden om indelingen niet al te absoluut te zien.

Statistische valstrikken

In product-research die meer fundamentele problemen tracht op te lossen, komt men vaak de volgende denkfout tegen; b.v. men tracht na te gaan of het publiek in het algemeen de voorkeur geeft aan ronde of vierkante druktoetsen op portable radio's. Men organiseert een test waarbij b.v. 200 potentiële kopers worden geconfronteerd met 2 portables: toestel A met ronde druktoetsen, toestel B met vierkante druktoetsen. De toestellen zijn verder identiek. Nu blijkt b.v. dat 90% de voorkeur geeft aan vierkante druktoetsen. Dit resultaat is stellig statistisch betekenisvol. Is nu bewezen dat het publiek in het algemeen de voorkeur geeft aan vierkante druktoetsen? Volstrekt niet! Deze test zou, statistisch bezien, veelvuldig herhaald moeten worden, steeds met andere typen portables. Daarenboven zouden we binnen het kader van deze tests ook moeten testen met steeds andere methodische opzetten zoals b.v. de splitruntechniek, m.a.w. afzonderlijke beoordelingen van de toestellen in onafhankelijke steekproeven. Pas als we bij voortduring een overwegende voorkeur voor vierkante druktoetsen zouden vinden, dus ongeacht de getoonde modellen, de omstandigheden en de gehanteerde onderzoeksmethoden, zouden we redelijkerwijze mogen aannemen dat we het fundamentele vraagstuk: 'ronde of vierkante druktoetsen' hebben opgelost . . . en dan nog alleen voor een beperkte tijd! We zouden dus om dit probleem op statistische wijze bevredigend op te kunnen lossen niet alleen een steekproef uit personen moeten trekken, maar ook steekproeven uit zowel modellen portables, als ook mogelijke onderzoekdesigns, omstandigheden en tijdperken (i.v.m. de invloed van mode). Dit is praktisch niet uitvoerbaar en zodoende zal men zich in de praktijk met één of enkele tests tevreden moeten stellen. Een waardevol hulpmiddel echter om deze tekorten te compenseren is het baseren van onze kennis

op de beantwoording van de vraag 'Waarom verkiest men toestel B wel/niet boven toestel A?' Als we de doorvraagprocedure naar de achtergronden van de keuze goed opzetten, krijgen we misschien gelegenheid te demonstreren dat vierkante drukknoppen onafhankelijk van allerlei andere factoren steeds, en met reden, worden verkozen boven ronde. *We hebben dan echter niet zo zeer meer met een streng wetenschappelijke als met een min of meer systematische 'gezond verstand' procedure te maken!*

Dit was een voorbeeld van een statistische valstrik m.b.t. aard en grootte van steekproeven. Een dergelijke fout wordt ook vaak gemaakt waar apparaten door het publiek op hun prestatie moeten worden beoordeeld. B.v. men laat maar één enkel scheerapparaat per type beoordelen door 100 mannen in een scheersalon. Men moet niet alleen een steekproef uit de massa van de beoordelaars trekken maar ook een steekproef uit de massa van alle apparaten van hetzelfde type, de z.g. 'bemonstering'.

De bovengenoemde problematiek kan zich ook in nog andere vormen van marktonderzoek voordoen, b.v. bij reclameresearch. Hierbij wordt b.v. gedacht aan een test die enige jaren geleden gehouden werd om na te gaan of het testen van het storyboard dezelfde resultaten oplevert als het testen van de gehele TV reclamefilm.

Dit werd nagegaan voor slechts twee onderwerpen en dit is nog niet voldoende om de resultaten statistisch te generaliseren.

Een andere statistische valstrik die niet direct tot het ABC van de marktonderzoeker behoort, kan optreden als men met een groot aantal items per interview werkt die achteraf op significantie getoetst worden.

'Op zich' is het gebruik van statistische toetsen al enigszins aanvechtbaar als de steekproef niet geheel 'at random' is getrokken, maar vermoedelijk is de fout niet heel erg groot als de steekproef op behoorlijke wijze tot stand is gekomen. Het uiteindelijk vrij willekeurige 5% (1 op 20) significantieniveau dat in marktonderzoekingen gebruikelijk is, zal b.v. misschien als gevolg van stratificatie enerzijds en clustersampling anderzijds in werkelijkheid kunnen liggen tussen 4% (1 op 25) en 7% (1 op 15). Erg is dit niet. Hoofdzaak is dat we een vast criterium (de toetsingsgrootheid) hebben om hypothesen al dan niet te verwerpen.

Erg begint het te worden als we dit criterium gebruiken om een hypothese te toetsen die pas gesteld is op het moment dat we het onderzoeksmateriaal voor het eerst inkeken (hypothese ad hoc!), een nog veel gevolgde praktijk helaas!

De hypothesen horen van te voren gesteld te zijn en daarbij dient tenminste vastgelegd te zijn of we tweezijdig of éénzijdig toetsen. Nog ernstiger is de zeer veelvuldig gemaakte fout om meer dan één item per subject *separaat* te toetsen en de uitkomsten op de geijkte manier te interpreteren. Dit is minder juist om de volgende redenen:

1. De mogelijkheid van testeffect is meestal onvoldoende uitgesloten (specifiek

volgorde effect van items op de nog volgende items, leerproces en vermoeidheid).

2. Zelfs als er geen sprake zou zijn van testeffect, zijn de gemeten items binnen één waarneming (interview) niet onafhankelijk van elkaar. De metingen vormen gewoonlijk een voor de geïnterviewde specifiek patroon.

In technieken als factoranalyse en clusteranalyse, tracht men de moeilijkheid die uit deze afhankelijkheid voortvloeit, te ondervangen door juist van deze afhankelijkheid uit te gaan. De items worden dan opgevat als diverse meetinstrumenten die alle 'hetzelfde' (factor, dimensie, vector, cluster) trachten te meten.

3. Ook bij veronderstelde volledige onafhankelijkheid van alle items wordt het optreden van een aantal 'spurious significances' waarschijnlijker naarmate het aantal items per subject toeneemt.

Uiteraard zonder het idee van 'een aantal items' per subject als zodanig te willen aanvallen, zou ik toch willen pleiten voor een veel voorzichtigere interpretatie van de resultaten van dergelijke waarnemingen dan meestal het geval is met materiaal dat niet speciaal voor factor-, clusteranalyse etc. is verzameld.

Maat-theoretische valstrikken

De moderne ontwikkelingen in de natuurkunde hebben ons denken aangaande het meten ingrijpend gewijzigd. Naarmate men overging tot het meten van steeds kleinere objecten (atomen, protonen, electronen, neutrino's), kwam meer en meer de moeilijkheid naar voren dat de activiteit van het meten een zodanig storende invloed op het te meten object uitoefent, dat dit daardoor in één of meer momentane eigenschappen gewijzigd wordt (b.v. snelheid, plaats, lading) en wel zo dat we de toestand op het moment van de meting zelf niet meer correct kunnen beschrijven. (We kennen dit verschijnsel onder de naam Heisenbergs onzekerheidsrelatie). Was voor de ontdekking van de onzekerheidsrelatie het meetresultaat een uitspraak over de eigenschappen van een gemeten object, thans is het veeleer slechts een uitspraak over de condities waaronder de meting verricht wordt. Het gemeten object is dan slechts één van deze condities. Als gevolg van deze gedachtengang zijn de natuurkundigen er tegenwoordig minder zeker van dat ze de werkelijkheid kennen door hun metingen en ze vatten de meetresultaten liever op als voorschriften (Doe dit, laat dat, meet verder op de volgende manier). In een dergelijke gedachtengang is een thermometer dus niet zozeer een ding waarop men kan zien hoeveel graden C. het ergens is als wel het belangrijkste, want regelend orgaan in een thermostaat. *In de sociale wetenschappen hebben we eveneens met de onzekerheidsrelatie te maken.* Het kleinste element dat we hier trachten te meten, is de individuele mens en ook deze kan door de meting (b.v. een interview) sterk in zijn momentane eigenschappen worden veranderd. Wij noemen dit verschijnsel 'interviewbias'. Ons streven moet er op ge-

richt zijn deze bias zo klein mogelijk te houden, waar dit kan te schatten en de resultaten zo veel mogelijk dynamisch te interpreteren in de zin van: 'doe dit! laat dat! meet verder op *deze* volgende manier!'.

De bevinding dat t.a.v. elementaire objecten (in ons geval de individuele mens) het meten zelf vaak meer tot het meetresultaat bijdraagt dan de gemeten objecten, verplicht ons ertoe nauwkeurig na te gaan wat de invloed is van:

a. *Het referentiekader, de achtergrond waartegen objecten worden gemeten*
Als we b.v. 'de attitude van Mr. X t.a.v. product A' willen meten (het tussen aanhalingstekens geplaatste is het te meten object), maakt het verschil of we alleen product A beschouwen (monadisch testen), een tweede product B in de vergelijking betrekken (paired comparison) of zelfs nog meer producten ter vergelijking aanbieden.

b. *De grootheden waarin de meting wordt uitgedrukt*
Ten aanzien hiervan kunnen we onderscheid maken tussen:
nominale grootheden, ook wel genoemd attributieve of kwalitatieve grootheden, b.v. de groothedenparen 0/1, +/–, ja/neen, de set attributen A.B.C.D.
ordinale of rangordenende grootheden
De grootheden zijn hier symbolen met een bepaalde volgorde b.v. de letters van het alfabet, de ranggetallen 1.2.3.4. etc. Een zeer gevaarlijke en veelvuldig voorkomende valstrik is de suggestieve werking die van de symboliek der ranggetallen uitgaat. Deze lijken nl. bedrieglijk veel op de natuurlijke getallen 1.2.3.4. etc., waarvoor b.v. geldt: '$2 \times 4 = 8$'. Voor ranggetallen geldt echter slechts: '$2<4<8$'.

Scalaire grootheden met interval karakter
Binnen dit systeem zijn uitspraken mogelijk van het karakter: '$4 - 2 = 6 - 4$', doch uitspraken van het type $2 \times 4 = 8$ zijn nog steeds niet mogelijk zoals wél het geval is met

Scalaire grootheden met ratio interval karakter
Binnen de scalaire grootheden wordt nog nader onderscheiden tussen continue en discontinue variabelen. Hiermee is het aantal mogelijkheden nog niet uitgeput (natuurlijke, rationele, complexe getallen, quaternions, tensoren) doch veel ervan hebben nog niet zo'n grote betekenis voor de sociale wetenschappen.

c. *De grootte en de begrenzing van de verzameling der meetgrootheden*
B.v. zijn nominale metingen binair (0,1) of b.v. quaternair (A, Q, X, S)? Werkt men met 5, 7, of 10 puntschalen? Gebruikt men ook negatieve waarden zoals in de schaal: '–3, –2, –1, 0, +1, +2, +3'?

Uit talrijke onderzoeken is komen vast te staan dat metingen aan eenzelfde object, doch uitgevoerd onder verschillende omstandigheden (verschillende re-

ferentiekaders, verschillende meetgrootheden), vaak zeer verschillende en niet tot elkaar te herleiden uitkomsten opleveren.

Twee belangrijke oorzaken voor deze discrepanties zijn:

a. dat men vaak een te eenvoudig systeem van formele relaties oplegt aan een veel gecompliceerder systeem van werkelijke relaties;

b. dat men vaak dubbelzinnige voorschriften geeft over de wijze waarop iemand zijn attitude moet scoren.

Voorbeeld van a

300 huisvrouwen wordt gevraagd uit 4 kleuren van overigens identieke koffiezetapparaten (lichtgrijs, blauwgrijs, donkergrijs en bruin) de aantrekkelijkste te kiezen. Resultaat:

(N = 300)	%	preferentievolgorde:
lichtgrijs	60	I
blauwgrijs	5	IV
donkergrijs	20	II
bruin	15	III

Als ook nog een tweede en derde voorkeur gevraagd wordt, zien we bij dezelfde mensen het volgende beeld:

(N = 300)	Keuze	1 %	2 %	3 %	4 %		preferentie volgorde:
lichtgrijs		60	20	15	5	100	I
blauwgrijs		5	70	10	15	100	II
donkergrijs		20	5	50	25	100	III
bruin		15	5	25	55	100	IV

We zien dat een *iets gecompliceerdere* wijze van meten al een heel ander resultaat oplevert. Verdere vragen zijn: wat zal de uitkomst zijn bij b.v. het gebruik van Stapelschalen? Is het wel juist de 4 kleuren tegen elkaar te testen als de fabrikant er uiteindelijk maar één uitbrengen wil? Drukt het woord 'preferentievolgorde' in beide gevallen hetzelfde uit?

Voorbeeld van b

Terwille van de overzichtelijkheid wordt dit voorbeeld slechts met 4 personen uitgewerkt. In werkelijkheid was er sprake van een sample van 80 personen.

richt zijn deze bias zo klein mogelijk te houden, waar dit kan te schatten en de resultaten zo veel mogelijk dynamisch te interpreteren in de zin van: 'doe dit! laat dat! meet verder op *deze* volgende manier!'.

De bevinding dat t.a.v. elementaire objecten (in ons geval de individuele mens) het meten zelf vaak meer tot het meetresultaat bijdraagt dan de gemeten objecten, verplicht ons ertoe nauwkeurig na te gaan wat de invloed is van:

a. *Het referentiekader, de achtergrond waartegen objecten worden gemeten*
Als we b.v. 'de attitude van Mr. X t.a.v. product A' willen meten (het tussen aanhalingstekens geplaatste is het te meten object), maakt het verschil of we alleen product A beschouwen (monadisch testen), een tweede product B in de vergelijking betrekken (paired comparison) of zelfs nog meer producten ter vergelijking aanbieden.

b. *De grootheden waarin de meting wordt uitgedrukt*
Ten aanzien hiervan kunnen we onderscheid maken tussen:
nominale grootheden, ook wel genoemd attributieve of kwalitatieve grootheden, b.v. de groothedenparen 0/1, +/–, ja/neen, de set attributen A.B.C.D.
ordinale of rangordenende grootheden
De grootheden zijn hier symbolen met een bepaalde volgorde b.v. de letters van het alfabet, de ranggetallen 1.2.3.4. etc. Een zeer gevaarlijke en veelvuldig voorkomende valstrik is de suggestieve werking die van de symboliek der ranggetallen uitgaat. Deze lijken nl. bedrieglijk veel op de natuurlijke getallen 1.2.3.4. etc., waarvoor b.v. geldt: '$2 \times 4 = 8$'. Voor ranggetallen geldt echter slechts: '$2<4<8$'.

Scalaire grootheden met interval karakter
Binnen dit systeem zijn uitspraken mogelijk van het karakter: '$4 - 2 = 6 - 4$', doch uitspraken van het type $2 \times 4 = 8$ zijn nog steeds niet mogelijk zoals wél het geval is met

Scalaire grootheden met ratio interval karakter
Binnen de scalaire grootheden wordt nog nader onderscheiden tussen continue en discontinue variabelen. Hiermee is het aantal mogelijkheden nog niet uitgeput (natuurlijke, rationele, complexe getallen, quaternions, tensoren) doch veel ervan hebben nog niet zo'n grote betekenis voor de sociale wetenschappen.

c. *De grootte en de begrenzing van de verzameling der meetgrootheden*
B.v. zijn nominale metingen binair (0,1) of b.v. quaternair (A, Q, X, S)? Werkt men met 5, 7, of 10 puntschalen? Gebruikt men ook negatieve waarden zoals in de schaal: '–3, –2, –1, 0, +1, +2, +3'?

Uit talrijke onderzoeken is komen vast te staan dat metingen aan eenzelfde object, doch uitgevoerd onder verschillende omstandigheden (verschillende re-

ferentiekaders, verschillende meetgrootheden), vaak zeer verschillende en niet tot elkaar te herleiden uitkomsten opleveren.

Twee belangrijke oorzaken voor deze discrepanties zijn:

a. dat men vaak een te eenvoudig systeem van formele relaties oplegt aan een veel gecompliceerder systeem van werkelijke relaties;

b. dat men vaak dubbelzinnige voorschriften geeft over de wijze waarop iemand zijn attitude moet scoren.

Voorbeeld van a

300 huisvrouwen wordt gevraagd uit 4 kleuren van overigens identieke koffiezetapparaten (lichtgrijs, blauwgrijs, donkergrijs en bruin) de aantrekkelijkste te kiezen. Resultaat:

(N = 300)	%	preferentievolgorde:
lichtgrijs	60	I
blauwgrijs	5	IV
donkergrijs	20	II
bruin	15	III

Als ook nog een tweede en derde voorkeur gevraagd wordt, zien we bij dezelfde mensen het volgende beeld:

(N = 300)	Keuze	1 %	2 %	3 %	4 %		preferentie volgorde:
lichtgrijs		60	20	15	5	100	I
blauwgrijs		5	70	10	15	100	II
donkergrijs		20	5	50	25	100	III
bruin		15	5	25	55	100	IV

We zien dat een *iets gecompliceerdere* wijze van meten al een heel ander resultaat oplevert. Verdere vragen zijn: wat zal de uitkomst zijn bij b.v. het gebruik van Stapelschalen? Is het wel juist de 4 kleuren tegen elkaar te testen als de fabrikant er uiteindelijk maar één uitbrengen wil? Drukt het woord 'preferentievolgorde' in beide gevallen hetzelfde uit?

Voorbeeld van b

Terwille van de overzichtelijkheid wordt dit voorbeeld slechts met 4 personen uitgewerkt. In werkelijkheid was er sprake van een sample van 80 personen.

4 proefpersonen, W, X, Y en Z, moeten 4 radio-portables, A, B, C en D, beoordelen op kwaliteit, volgens een puntensysteem van 1 t/m 7.

Portables:	Proefpersonen:				$\overline{M}$	Pref. volgorde:
	W	X	Y	Z		
A	4	6	4	5	4,75	II
B	1	4	5	6	4,00	IV
C	7	5	6	3	5,25	I
D	3	3	6	6	4,50	III

Wegens het geringe aantal (4) is hier i.p.v. de mediaan of modus het gemiddelde waarderingscijfer berekend; eigenlijk een aanvechtbare methode! Uit ondervraging achteraf is komen vast te staan dat p.p. het door hem als slechtst beoordeelde object een 1 heeft gegeven en het 'beste' een 7. X, Y en Z hebben blijkbaar bij de beoordeling een, ons onbekend, extern referentiekader gehanteerd, waardoor hun schalen eigenlijk van 3 tot en met 6 lopen. Stel dat we iedereen nu ondubbelzinnig verplichten het door W gevolgde systeem te volgen. We zouden dan krijgen:

Portables:	Proefpersonen:				$\overline{M}$	Pref. volgorde:
	W	X	Y	Z		
A	4	7	1	5	4,25	III
B	1	3	4	7	3,75	IV
C	7	5	7	1	5,—	I
D	3	1	7	7	4,50	II

We leren uit dit demonstratievoorbeeld het volgende:

– Door een wijziging in de instructie kan een verandering in de einduitslag optreden.

– Dit gevaar neemt toe naarmate het aantal schaalposities toeneemt.

Het gevaar dat een ondervraagde een schaal anders kan gebruiken dan voorzien is door de opsteller ervan (een maat-theoretisch vraagstuk), is van een zelfde orde als die van de gevaren die vooral een rol spelen bij de

Semantische valstrikken

Voor het ontstaan hiervan zijn vooral de volgende twee oorzaken van belang:

a. Het impliciete *geloof* dat er tussen een symbool (b.v. een woord) en de betekenis ervan een vaste relatie bestaat, terwijl in *werkelijkheid* tussen een symbool en een aantal betekenissen ($N \geqslant 1$) een min of meer losse relatie be-

staat die van persoon tot persoon kan verschillen en die ook in de tijd kan veranderen.

b. Het impliciete *geloof* dat een woord alleen naar een zaak verwijst (objectkarakter heeft), terwijl in *werkelijkheid* zeer veel woorden tegelijkertijd een attitude van de gebruiker tot uitdrukking brengen (subject karakter hebben).

Voorbeeld van a

Enige tijd geleden bleek uit een onderzoek dat het woord 'speciaalzaak' drager was van twee vrij sterk verschillende betekenissen die ieder bij ongeveer de helft van de ondervraagden voorkwamen.

Betekenis 1: een zaak waar men alleen artikelen van een vrij nauw omschreven soort kan kopen, b.v. een galanterieënwinkel, een aardappelhandel.

Betekenis 2: een zaak waar men een vrij breed assortiment artikelen verkoopt, maar alleen de luxere kwaliteiten.

Inmiddels zullen deze betekenissen naar inhoud en frekwentie binnen de spraakmakende gemeente alweer verschoven zijn. De semantische valstrik ontstond uiteraard onmiddellijk daar waar de onderzoeker zich niet bewust was van de mogelijkheid van dit onderscheid. Ook in de marketinggesprekken van de opdrachtgever was veel semantische verwarring over het woord 'speciaalzaak', evenals over het woord 'merkartikel':

Betekenis 1: een kwaliteitsartikel geleverd door een (meestal oude) zaak die gerenommeerd is in dit opzicht, b.v. Bussinks koek, de Gruyter's koffie.

Betekenis 2: een gewoon artikel dat een of andere merknaam op de verpakking draagt, b.v. de Gruyter's zeeppoeder.

Voorbeeld van b

Het bleek dat het veel verschil maakte of men ondervraagden verzocht hun preferentie aan te geven voor:

glimmende versus doffe koffiebonen

of voor:

glanzende versus doffe koffiebonen.

In een onderzoek met twee equivalente steekproeven waarin sprake was van 'glimmende' resp. 'glanzende' koffiebonen, bleek de voorkeur significant verschoven te zijn, naar de voorkeur voor glanzende koffiebonen indien dit woord gebruikt was. Daarentegen werden significant vaker doffe bonen geprefereerd als men het woord 'glimmend' als tegenstelling gebruikte. Blijkbaar brengt 'glimmend' een minder gunstige attitude tot uitdrukking dan 'glanzend' (vgl. een

glimmende neus, haar gelaat 'glansde' van vreugde maar 'glom' van het zweet), hoewel beide woorden naar een zelfde objectieve zaak verwijzen. Overigens moet gezegd worden dat een test als hierboven beschreven überhaupt niet valide is omdat het woordpaar glimmend (glanzend) - dof, niet voldoende is om alles van dofheid of glans uit te drukken. Men moet testen met monsters van echte koffiebonen.

Tot slot wordt nog de volgende mogelijk wat minder belangrijke semantische valstrik behandeld, nl.

c. Het impliciete *geloof*, dat woordparen die een tegengesteldheid uitdrukken ook verwijzen naar tegengestelde zaken, terwijl in *werkelijkheid* de betrokken zaken niet tegengesteld hoeven te zijn, en soms zelfs geheel geen verband met elkaar houden.

Voorbeeld van c

In de beginjaren van het psychologisch marktonderzoek werd in een test m.b.t. regenmantels gewerkt met een semantisch differentiaal, waarin o.a. voorkwam het termenpaar:

Voor degelijke vrouwen - voor niet zo degelijke vrouwen.

Deze term bleek in het geheel niet te differentiëren; zonder enige uitzondering bleken alle in het onderzoek betrokken regenmantels gezien te worden als 'voor degelijke vrouwen'. Een onderzoek naar de semantische waarde van het bovengenoemde termenpaar leerde het volgende:

Degelijke vrouwen: werden zeer overwegend beschreven als vrouwen die aan het huishouden een zekere zorg besteden (boel altijd netjes aan kant, goed voor de dag komen).

Niet zo degelijke vrouwen: werden zeer overwegend beschreven als vrouwen die het op zedelijk gebied niet zo nauw nemen (barvlinders, vrouwen die zich gewaagd kleden); een relatie met het huishouden kon niet gevonden worden.

De term: 'niet zo degelijk' die er 'op zich' als een understatement uitziet, was toenmaals zo pejoratief dat het niet te verwonderen viel dat vrijwel alle ondervraagden bij de scoring van het semantische differentiaal zo ver mogelijk uit de buurt van deze term bleven, ook al vonden zij de betrokken regenmantel niet zo bij 'degelijke vrouwen' passen. Vermoedelijk zouden thans de testresultaten weer anders uitvallen.

Slotconclusie

De hiervoren behandelde valstrikken vormen, hoewel veelvuldig voorkomend,

geenszins een uitputtende verzameling. Een conclusie die men dan ook zou kunnen trekken is dat voor het realiseren van een klassiek wetenschappelijke, d.w.z. strikt kwantitatieve aanpak van het produkttesten er nogal wat moeilijkheden onder het oog gezien moeten worden. Het lijkt dan ook niet aanbevelenswaardig om bij het streven hiernaar de 'kwalitatieve' methode van het gezond verstand los te laten. Hoe snel verliezen we immers niet onze mentale greep op exclusief kwantitatief materiaal dat ons van nature altijd ietwat onvertrouwd blijft en hoe licht zien we niet een goed verborgen valstrik over het hoofd. Uit het voorgaande kwam steeds naar voren dat door het toepassen van open interviewtechniek:

1. valstrikken hadden kunnen worden vermeden resp. achteraf zijn ontdekt en gecorrigeerd.
2. het mogelijk is toch min of meer de doelstellingen te realiseren van die onderzoekingen die in tijd en kosten niet haalbaar zouden blijken te zijn als men daarbij aan een exclusief kwantitatieve oplossing zou denken.

Het vergelijkend onderzoek van gebruiksgoederen bij de Consumentenbond

Ir. O. Grosch

Consumentenorganisaties zijn veelal ontstaan uit de behoefte van de verbruikers aan betere informatie, en speciaal bij de aanschaf van die goederen, waarvan de gecompliceerdheid de beoordeling van de kwaliteit vrijwel onmogelijk maakt.

De Nederlandse Consumentenbond werd in 1953 opgericht.

De bijdragen van de leden stellen de Consumentenbond in staat om vergelijkend onderzoek van waren en diensten uit te voeren en de resultaten te publiceren, individuele informatie te geven, en te bemiddelen of bijstand te verlenen bij geschillen.

Voor de uitvoering van al deze taken zijn ca. 50 man personeel in dienst, waarvan 10 uitsluitend ten behoeve van het vergelijkend warenonderzoek. Dit vergelijkend warenonderzoek beoogt naast het geven van produktinformatie aan de leden ook het stimuleren van kwaliteitsverbetering en het vergroten van prijsconcurrentie.

Plannenlijst

De aktiviteiten van de Consumentenbond komen het sterkst naar voren door de publikaties in de Consumentengids en door het werk van de afdeling ledenservice. Deze afdeling behandelt vragen en klachten van leden en is daardoor steeds goed geïnformeerd over de bij de leden levende wensen.

De meeste vragen betreffen duurzame huishoudelijke artikelen zoals wasmachines, stofzuigers en koelkasten, die dan ook regelmatig onderzocht moeten worden.

De keuze van de te onderzoeken artikelen wordt niet uitsluitend bepaald door de vragen van leden. Vanuit bestuur, directie, redactie en onderzoekstaf komen suggesties, waarmee soms op vragen van leden wordt geanticipeerd, soms geheel nieuwe groepen geïnteresseerden worden aangesproken. Publikaties in de

pers en alarmerende bevindingen bij buitenlandse consumentenorganisaties kunnen eveneens aanleiding zijn om een onderzoek te beginnen.

Voorts worden op gezette tijden onderzoeken herhaald, als het marktaanbod sterk gewijzigd is.

Tot nu toe zijn nog geen onderzoeken verricht naar gekwantificeerde gegevens omtrent de bij de leden levende wensen naar produktinformatie.

Door omvang en tijdsduur van het laboratoriumonderzoek moet de plannenlijst ca. een jaar vooruit vastliggen. Ze moet regelmatig worden aangevuld om een redelijke continuïteit in de Consumentengids te kunnen waarborgen.

Met consumentenorganisaties in naburige landen bestaat nauw kontakt. De mogelijkheid van gezamenlijk onderzoeken wordt regelmatig bekeken, en de plannenlijsten worden zonodig op elkaar afgestemd.

Overwegingen van stafbezetting, laboratoriumcapaciteit en seizoen bepalen verder uiteindelijk de keuze van de onderwerpen en de verdeling op de plannenlijst.

Testprogramma

De onderzoeker die met het onderwerp belast is begint een onderzoek met een marktverkenning. Hiertoe wordt fabrikanten of importeurs gevraagd om dokumentatie, worden advertenties bestudeerd en wordt nagegaan wat daadwerkelijk in een verscheidenheid van winkels verkrijgbaar is.

Om het laboratoriumonderzoek niet te langdurig of te kostbaar te maken, wordt dikwijls een keuze uit het totale aanbod gemaakt. Voor deze keuze kan een prijscriterium of een bepaalde uitvoeringsvorm worden gekozen (b.v. alleen wasmachines met wolwasprogramma). De voorgestelde keuzecriteria worden in het testprogramma omschreven. Soms wordt een groot onderzoek in tweeën gepubliceerd om per nummer tot een hanteerbaar aantal merken te komen (Cons.-gids jan. '69 naaimachines met vlakke plaat en febr. '69 naaimachines met vrije arm).

Om te voorkomen dat een produkt wordt onderzocht dat ingrijpende wijzigingen zal ondergaan of zelfs geheel van de markt zal verdwijnen, kan vooral bij langdurige onderzoeken van te voren over de keuze van het type met de fabrikant overleg nodig zijn.

De onderzoeker formuleert vervolgens de doelstelling en de uit te voeren tests. Hij raadpleegt daartoe deskundigen, gaat na welke eigenschappen van belang zijn (evt. door een vooronderzoek), verdiept zich in de vakliteratuur en de nationale of internationale normen.

Bij het opstellen van de testcriteria wordt ervan uitgegaan dat de gebruikswaarde van primair belang is, en niet de wijze waarop deze technisch is gereali-

seerd. De consument wil b.v. weten of een koelkast voldoende koelt, niet op welke wijze de koeling tot stand is gekomen.

Het technisch onderzoek wordt uitbesteed aan gespecialiseerde laboratoria in binnen- of buitenland. Veelal in samenwerking met de deskundigen van de laboratoria die offertes uitbrengen, worden de te onderzoeken gebruikseigenschappen vertaald in concrete metingen. De know-how van de onderzoekleider moet er voor waken dat in het laboratorium geen onnodige stokpaardjes van de vakspecialist worden bereden.

Naast laboratoriumonderzoek worden ook altijd gebruiksproeven uitgevoerd, waarbij het artikel gebruikt wordt op de wijze waarop een consument dat doet. Helaas is voor duurzame consumptiegoederen levensduuronderzoek slechts zelden mogelijk, omdat de onderzoektijd te beperkt is. Versnelde levensduurproeven bieden soms uitkomst, maar in veel gevallen is de relatie tussen deze proeven en de werkelijke gebruiksduur niet voldoende bekend.

Vergelijkend onderzoek

Een onderzoek naar de gebruikswaarde van een artikel verschilt principieel van statistische kwaliteitskontrole, concurrentieonderzoek of ontwikkelingsonderzoek. Bij statistische kwaliteitskontrole gaat het er veelal om spreiding, gemiddelde en uitvalspercentage van het fabrikageproces te bepalen. De consument echter is hierin niet geïnteresseerd. Hij koopt slechts één exemplaar van een artikel en verwacht hiervan bepaalde eigenschappen. Als deze geleverd worden is de koper tevreden, het gemiddelde laat hem koud. Defecten bij aankoop worden veelal door fabrikant of leverancier verholpen d.m.v. reparatie of ruil. Te geringe prestaties, die echter niet duidelijk aantoonbaar beneden de beloofde liggen, kunnen een gevolg zijn van fabrikagespreiding, maar ook van een verkeerde conceptie.

Produkten met een te geringe prestatie t.g.v. fabrikagespreiding hadden door doeltreffende kwaliteitskontrole niet op de markt hoeven te komen. Produkten met een verkeerde conceptie kunnen door de consument niet geruild worden voor een beter exemplaar van hetzelfde merk. Het vergelijkend onderzoek richt zich op de beoordeling van de algehele conceptie van het produkt, en niet op individuele meetuitkomsten, die spreiding kunnen vertonen. De algehele conceptie van het produkt bepaalt de gebruikseigenschappen.

Deze worden ontleed in meerdere deeleigenschappen. De individuele deeleigenschap heeft dus maar een beperkte invloed op de gebruikswaarde, en zo ook de spreiding in deze deeleigenschap. Zolang een minimum kwaliteit wordt bereikt of overschreden is de exacte meetwaarde niet zo interessant. De Consumentenbond onderzoekt daarom in principe slechts één exemplaar per merk.

Soms moeten echter wel individuele meetuitkomsten met spreiding worden beoordeeld. In dat geval worden meerdere exemplaren onderzocht en worden de regels van de statistiek toegepast om tot een verantwoorde presentatie te komen. Een voorbeeld hiervoor is het sigarettenonderzoek in de Consumentengids van febr. '69, waar teer- en nicotinegehalte in 95% waarschijnlijkheidsgebieden worden weergegeven.

Onderzoekmonsters, die door de inkopers van de Consumentenbond als een particuliere aanschaf worden ingekocht, worden eerst gecontroleerd op werking en op aanwezigheid van garantiebewijs en gebruiksaanwijzing en alsdan volgens testprogramma onderzocht.

Om zeker te zijn van de objectiviteit van de laboratoriumuitkomsten zal de onderzoeker waar mogelijk de monsters niet onder merk, maar onder codenummer aan het laboratorium beschikbaar stellen. Bij chemische analyses kan hij door het sturen van duplo's dan nog een indruk van meetonnauwkeurigheden krijgen.

Als het laboratoriumonderzoek en de gebruiksproeven beëindigd zijn worden de gegevens van belang voor de beoordeling van het produkt, per merk op een kontrolelijst verzameld en aan fabrikant of importeur van het onderzochte artikel toegestuurd ter commentaar. De binnenkomende reacties op de kontrolelijsten worden zorgvuldig bestudeerd. Deze kontrolelijsten bevatten alleen de gegevens van het eigen merk. De fabrikant krijgt dus niet de gegevens van de concurrent. Soms moet een meting worden herhaald ter kontrole of wordt over een bevinding met de fabrikant gesproken, waardoor wordt gewaakt tegen mogelijke vergissingen van laboratorium of onderzoeker, en uitschieters in de meetuitkomsten kunnen worden opgespoord.

Vergelijkende beoordeling

Uit de verkregen resultaten moeten conclusies worden getrokken. Meestal worden beoordelingen van belangrijke deeleigenschappen samengevat in een tabel, waarbij de klasseindeling: + + / + / □ / – / – – gehanteerd wordt.

De bepaling van de klassegrenzen en het trekken van conclusies gebeurt in overleg met de deskundigen. Om de verschillen tussen de merken tot hun recht te laten komen worden de klassegrenzen zodanig gekozen dat de resulterende beoordelingen vergelijkenderwijs ook een juist beeld geven. Omdat er uiteraard geen verschillen gesuggereerd moeten worden, die voor de gebruiker niet significant zijn, wordt bij de bepaling van de klassegrenzen daarmee rekening gehouden. De gebruiksproeven dienen hierbij als richtlijn.

De tabel met beoordeling per deeleigenschap stelt de lezer in staat zijn eigen keuze te doen. Zo wil b.v. de ene huisvrouw witter dan wit wassen, terwijl de

andere gelukkig is met de best centrifugerende machine en het waseffekt minder belangrijk vindt.

Om de veelheid van plussen en minnen te beperken wordt wel eens een afvalsysteem gebruikt. Bij het onderzoek van pechlampen (Cons.gids nov. '67) zijn b.v. alleen die lampen verder onderzocht op corrosie en windgevoeligheid die voldoende opvallend knipperlicht bezaten.

Het is ook mogelijk om door toekenning van wegingsfactoren aan de deeleigenschappen langs rekenkundige weg één eindwaardering per merk te bepalen. Deze methode, toegepast bij de beoordeling van meer dan honderd deeleigenschappen van kleurentelevisieontvangers (Cons.gids aug. '69), is echter niet zonder bezwaren.

De bepaling van de wegingsfactoren is slechts toe te passen wanneer de toekenning gemotiveerd kan worden.

Opgedane ervaringen

In de eerste jaren van zijn bestaan beschikte de Consumentenbond niet over de middelen om uitvoerige vergelijkende onderzoeken te verrichten.

Naarmate het aantal leden groeide, en daarmee ook de beschikbare gelden, konden er meer en grondiger onderzoekingen uitgevoerd worden. Op het ogenblik is een detaillist eigenlijk genoodzaakt om kennis te nemen van de Consumentengids, omdat zijn klanten anders beter georiënteerd kunnen zijn dan hijzelf. Deze situatie wordt niet altijd zonder protest geaccepteerd, omdat de vertrouwenspositie, die de winkelier graag bij zijn klanten wil bezitten, hierdoor wordt aangetast.

Er zijn echter ook winkeliers, die de produkteninformatie in de Consumentengids gebruiken bij de samenstelling van hun assortiment.

Door de publikatie in de Consumentengids zijn al herhaaldelijk (veelal noodgedwongen) artikelen verbeterd. Een voorbeeld daarvan zijn de reddingsvesten, waarvan bij het eerste onderzoek in juni '67 6 van de 50 onderzochte vesten als reddingsvest geschikt bleken. Na verloop van een jaar bleken zoveel veranderingen aangebracht te zijn dat bij een herhalingsonderzoek in mei '68 10 van de 13 merken voldeden.

De veelal verbreide mening dat er een positieve samenhang bestaat tussen prijs en kwaliteit wordt door de resultaten van het vergelijkend warenonderzoek effectief bestreden.

De produktinformatie uit een vergelijkend warenonderzoek is betrekkelijk snel na het verschijnen van de gids verouderd. De bij de bespreking naar voren gebrachte aspecten blijken veelal stimulerend te werken op het kritische beoordelen van een nieuw merk door de consument. Deze know-how blijft langer van waar-

de dan de produktinformatie en is daardoor een niet onbelangrijk educatief bijprodukt.

Slotbeschouwing

Niet altijd is het mogelijk om de beschreven procedure van het vergelijkend onderzoek volledig te volgen.

Gebrek aan bruikbare onderzoekmethoden of laboratoria, aan tijd of aan middelen kunnen de onderzoekleider doen besluiten om een deel van het onderzoekprogramma niet uit te laten voeren. Zolang het ondergeschikte aspecten betreft, en door de weglating het onderzoek niet aan voorlichtende waarde inboet, is deze wijze van uitvoering aanvaardbaar.

De methodiek van het vergelijkend onderzoek is nog jong en sterk in ontwikkeling.

Niettemin is uit de tot nu toe opgedane ervaringen reeds gebleken dat grote kwaliteitsverschillen verdwijnen.

Misschien maakt het vergelijkend warenonderzoek zichzelf op den duur weer overbodig.

Literatuur

DIN 66050 Gebrauchtauglichkeit, Begriff
DIN 66051 Untersuchung von Waren
Consumentengids aug. '68: Hoe pakt de Consumentenbond de zaken aan?

Beoordeling van de kwaliteit van droogscheerapparaten

P. Camfferman, C. P. L. Commissaris en A. W. J. de Gee

Inleiding

Een electrisch droogscheerapparaat is een typische vertegenwoordiger van een betrekkelijk kleine groep producten, waarvoor het vaststellen van een objectief oordeel over de kwaliteit door middel van enquêtering van proefpersonen op bijzondere moeilijkheden stuit, ook wanneer slechts een relatief kwaliteitsoordeel wordt verlangd.

Men heeft immers te maken met een product dat enerzijds door zijn uiterlijke verschijning zijn herkomst onmiddellijk verraadt, terwijl anderzijds de kwaliteit wordt bepaald door moeilijk definieerbare begrippen als 'scheerkwaliteit' en 'huidirritatie', ten aanzien waarvan de waardering zeer sterk afhankelijk is van de 'huid- en haareigenschappen' van de gebruikers. Soortgelijke problemen doen zich, ofschoon in mindere mate, voor bij andere 'personal care' artikelen als electrische tandenborstels, massage-apparaten e.d. Toch is het voor fabrikanten en consumenten van belang dat een methode wordt ontwikkeld, die de formulering van een zo objectief mogelijk kwaliteitsoordeel toestaat. Beide groepen belanghebbenden immers wensen te beschikken over gegevens die het mogelijk maken de 'impact' van de intensieve reclamecampagnes, die vooral de verkoop van 'personal care' artikelen begeleiden, te scheiden van de invloed van de feitelijke kwaliteit van de producten.

Men kan zich afvragen of het niet mogelijk is werkelijk objectieve methoden tot kwaliteitsbeoordeling van droogscheerapparaten te ontwikkelen. In dit verband kan men b.v. denken aan bepaling van de lengteverdeling van de haarstoppels, die na het scheren op het gelaat overblijven, aan meting van de tijdens het scheren geproduceerde hoeveelheid huidschilfers, en, eventueel, aan meting van de 'doorbloeding' van de huid ten gevolge van schrijnen. Afgezien van het feit dat de ontwikkeling van dergelijke methoden de technicus voor bijzonder moeilijke problemen stelt, geldt dat bij de beoordeling van de eindkwaliteit een aantal subjectieve zaken als 'hanteerbaarheid', 'vorm' (in verband met de bereikbaarheid van de moeilijk bereikbare plaatsen van het gelaat), e.d. een wezenlijke rol meespelen. Uiteraard dient een goede methode de invloed van de voor

elke functie van het apparaat niet wezenlijke eigenschappen, als b.v. kleur en prijs, te elimineren. Hieruit volgt dat men ten behoeve van een volledige voorlichting van fabrikant en consument subjectieve kwaliteitsbeoordeling door middel van enquêtering niet kan missen.

In het hiernavolgende zullen wij nagaan welke faktoren primair van belang kunnen worden geacht voor de kwaliteit van een droogscheerapparaat en hoe men door enquêtering tot een zo objectief mogelijk kwaliteitsoordeel kan geraken. Aan de hand van een aantal voorbeelden zal worden aangetoond welke maatregelen men moet nemen om misleidende uitspraken te vermijden.

Opzet en uitvoering van het onderzoek

Beoordelingscriteria

Men kan van mening verschillen over de criteria, die bepalend zijn voor de kwaliteit van een droogscheerapparaat, en stellig over de relatieve 'zwaarte' van elk dezer criteria.

Ofschoon een nog uit te voeren, separate enquêtering zal moeten uitwijzen of de 'hoofdcriteria' juist zijn gekozen, werd het volgende lijstje van 'hoofdcriteria' opgesteld[1]:

1. Uiteindelijk scheerresultaat (samengesteld uit 'eindgladheid', 'gelijkmatigheid', 'bereikbaarheid van moeilijk te bereiken plaatsen' en 'gemak waarmede lange haren worden geschoren');
2. Scheersnelheid (samengesteld uit 'totaal indruk ten aanzien van de snelheid' en 'gladheid na 1 minuut scheren');
3. Huidirritatie;
4. Hanteerbaarheid van het apparaat tijdens het scheren;
5. Neveneffekten (temperatuur, geluid en trillingen);
6. Samenvattend oordeel betreffende diverse ergonomische eigenschappen als 'gemak van reinigen', 'opruimen', e.d.

Daarnaast geldt voor batterij- of accu-apparaten de gebruiksduur van de batterijen, respectievelijk de duur gedurende welke een eenmaal opgeladen apparaat kan worden gebruikt, als kwaliteitscriterium. Ook de eventuele aanwezigheid en kwaliteit van accessoires (b.v. tondeuze) kunnen van invloed zijn op het eindoordeel.

Zoals bekend tenderen sommigen tot het geven van een samenvattend eindoordeel. Hiertoe worden de bovengenoemde 'hoofdeigenschappen' van een 'gewichtsfaktor' voorzien, waarna een 'gewogen' gemiddeld eindoordeel wordt berekend. Uiteraard is een dergelijke procedure volstrekt zinloos wanneer het kwaliteitsonderzoek wordt uitgevoerd ten behoeve van de fabrikant, ofschoon kennis van de gewichtsfaktoren als zodanig van commercieel belang is. Aangezien wij

van mening zijn dat men de mondigheid van de consument niet moet onderschatten, moet men onzes inziens eveneens afzien van het geven van een gewogen gemiddeld eindoordeel, wanneer het doel van het onderzoek is de consument voor te lichten over de relatieve kwaliteit van een aantal scheerapparaten. Men gaat wel zeer ver over de schreef wanneer men een dergelijk eindoordeel bovendien nog combineert met de prijs.

Met uitzondering van de punten 'huidirritatie' en 'hanteerbaarheid', zijn de bovengenoemde hoofdcriteria samengesteld uit een aantal detailcriteria. Uiteraard dient de fabrikant ten behoeve van de voortgezette ontwikkeling van zijn apparaat alle mogelijke detail-informatie te krijgen die een verstandig opgezet onderzoek redelijkerwijs kan verschaffen. Anderzijds mag men het de consument niet lastig maken met een te zeer gedetailleerde informatie. Het 'wegen' van detail-criteria is dus een onvermijdelijke stap bij het formuleren van een eindoordeel voor ieder der samengestelde hoofdcriteria. Zo mogelijk dient de relatieve 'zwaarte' van elk der detail-criteria te zijn gebaseerd op statistisch onaanvechtbaar cijfermateriaal, b.v. verkregen door enquêtering van een grote, a-selecte groep proefpersonen.

Vragenlijst en scoringsmogelijkheden

Nadat men het eens is geworden over de beoordelingscriteria kan een vragenlijst worden opgesteld. Deze moet, met behoud van de voor enquêtering noodzakelijke eenvoud, het maximum aan informatie verschaffen. De vragen die thans bij enquêtering in Nederland worden gesteld, zijn de volgende:

1. Hoe lang heeft U zich met het apparaat moeten scheren?
2. Hoe beoordeelt U de bereikte gladheid na een minuut scheren?
3. Hoe beoordeelt U de eindgladheid?
4. Hoe beoordeelt U de gelijkmatigheid van het scheerresultaat?
5. Hoe beoordeelt U de bereikbaarheid van de moeilijk te bereiken delen van het gelaat?
6. Hoe beoordeelt U het gemak waarmee lange haren worden afgeschoren?
7. Veroorzaakt het scheerapparaat huidirritatie?
8. Hoe ligt het apparaat in de hand tijdens het scheren?
9. Hoe ervaart U de temperatuur van de scheerkop?
10. Hoe ervaart U het geluid dat het scheerapparaat tijdens het scheren maakt?
11. Hoe ondervindt U de trillingen die tijdens het scheren in het apparaat optreden?
12. Hoe beoordeelt U de reinigingsmogelijkheid van het apparaat?
13. Hoe beoordeelt U het gemak waarmee het apparaat in de doos kan worden opgeborgen?
14. Hoe beoordeelt U de totale kwaliteit van het apparaat?

Alleen voor oplaadbare apparaten:

15. Hoe beoordeelt U het gemak waarmee het apparaat kan worden opgeladen?

16. Hoeveel maal heeft U zich met het scheerapparaat kunnen scheren?

Voor de wijze waarop de beantwoording van de bovenstaande vragen mogelijk wordt gemaakt bestaan verschillende systemen, waarbij men in eerste instantie een keuze moet doen tussen 'analoge' en 'numerieke' systemen van scoring. Een vaak gehanteerd analoog systeem is de 'thermometerschaal', bestaande uit een verticale balk waarop de proefpersoon door middel van een kruisje zijn appreciatie moet aangeven. Bij numerieke beoordeling moet men een principiële keuze maken tussen systemen waarbij wordt becijferd van 1 tot 10 (of van 1 tot 5) en systemen waarbij wordt becijferd rond 0, dus bijvoorbeeld van –3 tot +3.

Een aanpassing aan nationale situaties bij toepassing van numerieke systemen lijkt geboden. In het algemeen echter moet bij toepassing van numerieke systemen voorzichtigheid worden betracht. Een bekend Nederlands numeriek systeem 1-10 (1-5 is onvoldoende, 6-10 is voldoende) heeft door de praktijk op scholen zozeer geleden ten aanzien van het normbesef, dat het nauwelijks nog als lineaire schaal bruikbaar is. Hoezeer dit juist is, blijkt uit het tijdens onze onderzoeken toegepaste +3/–3 systeem, waarbij bleek dat er een aversie bestaat tegen het gebruik van de nul.

Veel bedenkelijker is uiteraard de niet-lineaire schaaltransformatie, die ontstaat door proefpersoonuitspraken als: goed, middelmatig, slecht, via 3, 2, 1 om te zetten in 7, 3, 1 om daardoor voldoend grote verschillen in uitspraken te bereiken.

In het huidige stadium van onderzoek staat het voor ons nog geenszins vast welke scoringsmogelijkheid de voorkeur verdient, alleen al omdat de wijze waarop het onderzoek wordt uitgevoerd mede bepalend is voor de keuze van het scoringssysteem. Tenslotte zal de praktijk moeten uitwijzen welk systeem in een bepaald geval de voorkeur verdient. Ofschoon wij niet gaarne de indruk zouden wekken dat wij nu reeds een eindoordeel over de optimale scoringsmethode zouden kunnen formuleren, beperken wij ons in het hierna volgende tot resultaten, behaald door toepassing van het '+3/–3 systeem.

Grootte en samenstelling van het panel

Een belangrijke vraag bij het opstellen van een schema voor een subjectief kwaliteitsonderzoek is die naar het aantal proefpersonen dat bij het onderzoek wordt ingeschakeld. Daarbij moet men in aanmerking nemen dat spreiding in de resultaten wordt veroorzaakt door de volgende faktoren:

a. de spreiding in de eigenschap ten gevolge van productiesysteem;

b. de spreiding in de behoeften van de consument ten aanzien van de onderhavige eigenschap;

c. het vermogen tot beoordeling van een bepaalde eigenschap door de proefpersoon.

In het algemeen stijgt met het aantal proefpersonen het 'scheidend vermogen' van de methode tot een asymptotische waarde is bereikt. Dit is een gevolg van het feit dat een relatief kwaliteitsverschil tussen twee typen droogscheerapparaten slechts als zodanig wordt onderkend, wanneer dit verschil statistisch significant van nul afwijkt. Anderzijds wordt de grootte van het panel beperkt door de eis dat de kosten, waarmee het onderzoek gepaard gaat, niet te hoog mogen oplopen. Men zou de afmeting van het panel kunnen optimaliseren door een onderzoek van verschillende typen droogscheerapparaten herhaalde malen uit te voeren, met steeds toenemend aantal proefpersonen. Op het moment dat een voortgezette toename van het aantal proefpersonen geen verdere, statistisch significante informatie meer oplevert, is de optimale panelgrootte bereikt. Het behoeft geen betoog dat een dergelijke werkwijze buitengewoon tijdrovend is en daardoor vrijwel niet te financieren. In het hiernavolgende worden resultaten besproken van een onderzoek, dat werd uitgevoerd met medewerking van twee panels van 24 personen elk. Het zal blijken dat deze panels in elk geval groot genoeg waren om tot een representatief en reproduceerbaar oordeel te geraken. Wat samenstelling betreft moet elk panel aan de volgende elementaire eisen voldoen:

1. het panel moet een redelijk representatieve doorsnede vormen van de mannelijke bevolking tussen 20 en 60 jaar, waarbij speciaal moet worden gelet op baard- en huideigenschappen;

2. het panel moet zodanig zijn samengesteld dat elk te onderzoeken merk door een gelijk aantal proefpersonen in het panel is vertegenwoordigd. Wij zullen aan de hand van resultaten aantonen dat deze eis volstrekt essentieel is, wanneer men wil komen tot een reële beoordeling van de kwaliteit van droogscheerapparaten.

Aantal apparaten

Een andere belangrijke vraag bij de opstelling van een onderzoeksschema voor subjectieve kwaliteitsbeoordeling is die naar het aantal te onderzoeken apparaten. Er wordt nogal eens aangenomen dat onderzoek van slechts één apparaat per type voldoende informatie zou verschaffen om tot een verantwoord oordeel over de relatieve kwaliteit van verschillende typen droogscheerapparaten te geraken, en wel op grond van de volgende overwegingen:

– een consument koopt in het algemeen ook slechts één apparaat

– elk apparaat dat te koop is moet goed zijn.

Ofschoon wij ons van harte met de tweede overweging kunnen verenigen, be-

strijden wij met grote stelligheid de gedachte dat men een verantwoord oordeel zou kunnen baseren op de aankoop van slechts één apparaat per type en dit wel op grond van een overweging die is te beschouwen als een inversie van de eerstgenoemde. Juist immers omdàt een consument zich niet kan permitteren meer dan één apparaat te kopen, moet hij worden 'beschermd' tegen eventueel optredende variaties in kwaliteit. Dit standpunt wordt verduidelijkt in Figuur 1 waarin twee (hypothetische) kwaliteitsverdelingscurven zijn getekend. Men ziet dat de 'mediaankwaliteit' van product I aanmerkelijk beter is dan die van product II, terwijl tevens de kwaliteitsbeheersing in het geval van product I aanmerkelijk beter is dan die voor product II. Toch kan het, door onfortuinlijke bemonstering, voorkomen dat een droogscheerapparaat van type II wordt aangeschaft dat van betere kwaliteit is dan het aangeschafte apparaat van type I. Een dergelijke situatie dient men uiteraard tot elke prijs te vermijden.

Het aantal apparaten dat men minimaal moet bemonsteren om tot een verantwoorde uitspraak omtrent de kwaliteit te geraken is niet op theoretische gronden af te leiden. De onderzoeken, die tot op heden door ons werden uitgevoerd, werden verricht aan 15 apparaten per type. Alleen een herhaalde malen uitgevoerd onderzoek, steeds met verschillende aantallen apparaten per type zou kunnen uitwijzen of dit aantal optimaal is.

Wijze van uitgifte en feitelijke uitvoering

Wij beëindigen deze bespreking van de methode met te wijzen op de noodzaak dat de proefpersonen zich onder normale gebruiksomstandigheden, dat wil zeggen thuis, met de apparaten moeten scheren. Gecombineerd met de voor de hand liggende voorwaarde dat de proefpersonen volstrekt onafhankelijk van elkaar tot een oordeel moeten geraken, en elkaar dus niet mogen kennen, leidt dit tot een situatie die zo goed mogelijk representatief is voor de praktijk van het droogscheren. Omdat men enige tijd nodig heeft om zich te gewennen aan een nieuw type scheerapparaat, dient de uitgifteperiode niet te kort te zijn. De ervaring heeft ons geleerd dat een uitgifteperiode van 14 dagen voor de meeste praktisch voorkomende gevallen lang genoeg is. Ook in dit geval geldt echter weer dat een verantwoorde uitspraak over de minimaal noodzakelijke duur van de uitgifteperiode pas mogelijk zou zijn na uitgebreid onderzoek.

Tenslotte kan men zich afvragen of de scheerapparaten paarsgewijs of achtereenvolgens moeten worden verstrekt. Bij paarsgewijze uitgifte wordt elk oordeel *relatief* ten opzichte van het andere type apparaat, dat tegelijk werd uitgegeven, geformuleerd. Dit is in elk geval een essentiële voorwaarde wanneer men gebruik maakt van een 'analoge' scoringsmethode. Verder wordt bij paarsgewijze uitgifte een eindoordeel verkregen dat onafhankelijk is van eventuele graduele verschuivingen in de beoordelingsmaatstaf gedurende het onderzoek. Een duidelijk nadeel van paarsgewijze uitgifte is echter dat de duur van het onderzoek bij

toenemend aantal typen snel stijgt, wanneer men de voor de hand liggende eis stelt dat elke proefpersoon elke mogelijke combinatie van apparaten minstens éénmaal moet hebben beoordeeld. Wanneer men uitgaat van n typen apparaten, bedraagt het aantal mogelijk te maken ongelijke combinaties

$$\tfrac{1}{2}\,n(n-1)$$

Hieruit volgt dat bij gebruik van één panel, onderzoek van vijf verschillende typen, bij een uitgifteperiode van 14 dagen per combinatie, reeds minstens 20 weken duurt. Dit kan een groot bezwaar zijn wanneer men een zo groot mogelijk percentage van de op de markt zijnde apparaten in het onderzoek wil betrekken[2]. Afgezien van de financiële consequenties verliest een onderzoek immers aan slagvaardigheid wanneer het te lang duurt! In verband met deze overweging werd één bepaald onderzoek, waarbij vier typen droogscheerapparaten met elkaar werden vergeleken, tweemaal uitgevoerd, te weten éénmaal met paarsgewijze uitgifte en éénmaal met achtereenvolgende uitgifte. Daarbij werd in beide gevallen gebruik gemaakt van het '+3/–3' scoringssysteem. Enigszins tegen de aanvankelijke verwachting in, wijzen de resultaten uit dat achtereenvolgende uitgifte geen nadelige invloed behoeft te hebben op de scherpte van de uitspraken ten aanzien van de relatieve kwaliteit van de apparaten.

Enkele resultaten

Gebaseerd op de in het voorgaande gepresenteerde beschouwingen, werd met medewerking van de Nederlandse Stichting voor de Statistiek (NSVS) te Den Haag, een onderzoek uitgevoerd naar de relatieve kwaliteit van zeven verschillende typen droogscheerapparaten. Uiteraard staat het bestek van deze voordracht ons niet toe een volledig overzicht van de resultaten te presenteren. Wij willen echter trachten de juistheid van een aantal eerder gedane beweringen te bewijzen door presentatie van enkele representatieve resultaten. In het bijzonder zullen wij trachten de volgende vragen te beantwoorden:

1. Is een panel van 24 proefpersonen groot genoeg om een representatieve uitspraak ten aanzien van de relatieve kwaliteit van een scheerapparaat te verkrijgen?
2. Hangt de waardering die aan een bepaald type apparaat wordt toegekend al dan niet af van de kwaliteit van de overige apparaten die aan de proefpersoon ter beoordeling worden voorgelegd, of kan een kwaliteitsoordeel als 'absoluut' worden gezien?
3. Is het geoorloofd om meerdere typen apparaten van één fabrikaat aan één en hetzelfde panel voor te leggen?

4. Is het werkelijk noodzakelijk dat in elk panel elk 'fabrikaat' in gelijke mate is vertegenwoordigd?

Zoals gezegd had het onderzoek betrekking op een onderlinge vergelijking van zeven typen droogscheerapparaten. Deze werden gefabriceerd door drie verschillende fabrikanten: A, B en C. In het hiernavolgende worden de zeven typen gekarakteriseerd door de cijfers 1 t/m 7. Het onderzoek was gebaseerd op paarsgewijze uitgifte en het '+3/–3' scoringssysteem werd toegepast. Teneinde na te gaan of een panel van 24 proefpersonen een representatief en reproduceerbaar oordeel uitspreekt, werden twee onafhankelijke panels gevormd, volgens de hiervoor genoemde eisen. Dit houdt onder meer in dat in elk panel acht 'A scheerders', acht 'B scheerders' en acht 'C scheerders' voorkwamen. Teneinde na te gaan in hoeverre subjectieve kwaliteitsbeoordeling door enquêtering een 'absoluut' waardeoordeel mogelijk maakt, werden de apparaten in twee groepen gesplitst, en wel zodanig dat de eerste groep werd gevormd door de apparaten 1, 2, 3 en 4 en de tweede groep door de apparaten 4, 5, 6 en 7. Apparaat 4 kwam dus in beide groepen voor en diende als 'referentie-apparaat'. Het schema dat men vindt in Figuur 2 geeft een beeld van de wijze waarop het onderzoek werd uitgevoerd. Tijdens periode 1, die zich uitstrekte van 1 mei tot 14 juni 1968, werkte panel I mee aan het onderzoek van apparatengroep I, terwijl panel II zich bezighield met apparatengroep II. Gedurende de tweede periode, die zich uitstrekte van 8 juli tot 23 augustus 1968, beoordeelde panel II apparatengroep I, terwijl de apparaten uit groep II aan panel I werden voorgelegd.

Een voorbeeld voor het soort resultaten dat men verkrijgt, vindt men in Tabel 1, waarin men de gemiddelde scores ziet voor de apparaten 1, 2, 3 en 4, berekend op basis van de individuele scores, gegeven door panel II gedurende de tweede periode van onderzoek, voor zover deze betrekking hebben op de vragen 3 (totale eindgladheid), 5 (bereikbaarheid), 6 (gemak waarmee lange haren worden geschoren) en 7 (huidirritatie). In deze tabel zijn de apparaten gerangschikt in dalende volgorde van kwaliteit. Teneinde na te gaan of de optredende verschillen tussen de gemiddelde scores onderling statistisch significant verschillen, werd een analyse volgens Duncan uitgevoerd. Een beschrijving van deze analyse vindt men in Ref. 1. De resultaten worden in de tabel aangegeven door middel van onder de getallen getrokken strepen, waarbij een doorlopende streep onder 2, 3 of 4 getallen aanduidt dat deze tot één statistisch homogene groep behoren. Wanneer dit het geval is, mogen geen conclusies worden getrokken ten aanzien van de onderlinge rangorde van de desbetreffende resultaten. De 'kwaliteit' van de betrokken groep kan worden gekarakteriseerd door opgave van de gemiddelde score. Uit de resultaten, vermeld in Tabel 1, kunnen dus de volgende, statistisch verantwoorde, conclusies worden getrokken:

1. De apparaten 1, 2 en 4 scheren gladder dan apparaat 3. Ten aanzien van

eventuele verschillen tussen de apparaten 1, 2 en 4 onderling mogen geen conclusies worden getrokken.

2. De moeilijk te bereiken plaatsen op het gelaat zijn gemakkelijker te bereiken met apparaat 2 dan met de andere drie apparaten. Wederom mogen geen conclusies worden getrokken ten aanzien van eventuele onderlinge verschillen tussen de apparaten 1, 3 en 4.

3. Apparaat 1 veroorzaakt meer huidirritatie dan de apparaten 2, 3 en 4. Ten aanzien van eventuele verschillen tussen de laatstgenoemde apparaten onderling zijn geen uitspraken mogelijk.

4. Ten aanzien van het gemak waarmede lange haren kunnen worden afgeschoren, kan slechts worden geconcludeerd dat apparaat 1 dit beter doet dan apparaat 3. Omtrent de relatieve positie van de apparaten 2 en 4 is geen andere uitspraak mogelijk dan dat zij in elk geval niet beter zijn dan apparaat 1 en niet slechter zijn dan apparaat 3.

Het bovenstaande wijst uit dat men nogal eens met een aantal vaagheden blijft zitten wanneer men zich beperkt tot de statistisch verantwoorde conclusies. Meer informatie kan alleen worden verkregen door uitbreiding van het onderzoek met andere proefpersonen. Vaak zullen daartoe tijd en financiële middelen ontbreken. Wellicht ten overvloede waarschuwen wij met klem tegen het gevaar voor onjuiste uitspraken, dat ontstaat wanneer men desondanks meent meer gedetailleerde conclusies te moeten produceren door negeren van de statistiek!

Teneinde de vragen 1 t/m 4 te kunnen beantwoorden, berekenden wij ten eerste voor alle onderzochte apparatentypen, de totaal gemiddelde scores die betrekking hebben op het scheerresultaat. Hiertoe werden, *zonder rekening te houden met eventueel verschil in 'gewicht'*, de individuele scores die betrekking hebben op de vragen 1 (eindgladheid), 2 (gelijkmatigheid), 3 (bereikbaarheid van moeilijk bereikbare plaatsen) en 4 (gemak waarmee lange haren worden afgeschoren) gemiddeld. De resultaten vindt men in Tabel 2.

Uit deze tabel blijkt dat panels I en II het bijzonder goed eens zijn over de scheerkwaliteit van de apparaten 1, 2, 3, 5, 6 en 7. Hieruit volgt dat met een groep van 24 personen een representatief en goed reproduceerbaar scoregemiddelde kan worden verkregen. Hiermede is vraag 1 bevestigend beantwoord.

De resultaten behaald met apparaat 4 lijken geheel in strijd te zijn met deze conclusie. Van overeenstemming tussen de beide panels is in dit geval namelijk geen sprake. Wel blijkt, dat de panels elk voor zich zeer 'standvastig' zijn in hun oordeel. De verklaring voor dit opmerkelijke verschijnsel is ongetwijfeld dat de proefpersonen van panel I aanvankelijk geconfronteerd werden met apparaat 4 in gezelschap van de relatief 'minder goede' apparaten 1, 2 en 3. De relatief goede indruk die men in verband daarmee van apparaat 4 kreeg, is men kennelijk niet kwijt geraakt tijdens de beoordeling van de apparaten van groep II tijdens perio-

de 2. De proefpersonen van panel II daarentegen kregen aanvankelijk apparaat 4 ter beoordeling in een groep waarin zich ook de apparaten 5, 6 en 7 bevonden. Kennelijk houdt apparaat 4 zich in gezelschap van deze apparaten minder goed. Ook deze 'slechte' indruk herinnert men zich, hetgeen blijkt uit de relatief lage score die de proefpersonen van panel II toekennen aan apparaat 4, wanneer dit tijdens de tweede periode wordt vergeleken met de apparaten 1, 2 en 3. Hieruit blijkt zeer duidelijk dat men bij een subjectief kwaliteitsonderzoek slechts *relatieve* informatie verkrijgt, hetgeen wil zeggen dat men slechts te weten komt hoe een bepaald type droogscheerapparaat zich verhoudt tot een aantal andere typen waarmee het desbetreffende apparaat 'toevallig' wordt vergeleken. In gezelschap van relatief 'slechte' apparaten wordt een bepaald type 'goed' beoordeeld, in gezelschap van relatief 'goede' apparaten wordt hetzelfde type 'slecht' beoordeeld. Dit bewijst dat men nooit of te nimmer aan de resultaten van subjectief kwaliteitsonderzoek, verkregen door enquêtering van proefpersonen, absolute waarde mag toekennen. Hiermee is vraag 2 beantwoord.

Kennelijk herinnert men zich een 'goede', respectievelijk een 'slechte' indruk zeer nauwkeurig tijdens de uitvoering van een volgend onderzoek. Aan het optreden van een dergelijk 'herinneringseffekt' moet men de conclusie verbinden dat men eenzelfde apparaat niet tweemaal aan hetzelfde panel proefpersonen ter beoordeling mag voorleggen, in elk geval niet dan nadat er een voldoend lange 'ontwenningsperiode' in acht is genomen. Overigens is het zeer waarschijnlijk dat het 'herinneringseffekt' zich ook manifesteert wanneer twee op elkaar gelijkende typen apparaten van één fabrikaat aan eenzelfde panel ter beoordeling worden aangeboden. In verband hiermee zijn wij van mening dat vraag 3 ontkennend moet worden beantwoord.

Wij achtten het buitengewoon interessant om na te gaan of de geconstateerde invloed van de andere apparaten met het bijbehorend 'herinneringseffekt' zich bij alle drie de soorten proefpersonen voordoet. Hiertoe werden de resultaten, die betrekking hebben op beide onderzoeksperioden, uitgesplitst per soort proefpersonen. De resultaten die betrekking hebben op de beoordeling van apparaat 4 vindt men in Tabel 3. Opnieuw blijkt een belangwekkend feit! De geconstateerde groepsinvloed met bijbehorend herinneringseffekt manifesteert zich zeer sterk bij de 'A scheerders' en de 'B scheerders'. Zeer opmerkelijk is echter dat de gemiddelde resultaten, gescored door de 'C scheerders' één, statistisch homogene groep vormen. De verklaring voor het feit dat het oordeel dat de 'C scheerders' vellen over type 4 niet wordt beïnvloed door de kwaliteit van de overige typen is dat de 'C scheerders' ten aanzien van apparaat 4 (dat immers een 'C apparaat' is!) zozeer zijn 'gepredestineerd', dat het geven van een onafhankelijk, vergelijkend oordeel onmogelijk is geworden. Anders gezegd: wanneer 'C scheerders' worden geconsulteerd, staat de gemiddelde score voor apparaat 4 vast, ongeacht de apparaten waarmee apparaat 4 wordt vergeleken. Hieruit kan reeds met zeker-

heid worden afgeleid dat het eindoordeel van proefpersonen in zeer sterke mate wordt bepaald door hun aan het onderzoek voorafgaande scheergewoonten.

Dit blijkt ook wanneer men de antwoorden op de enquêtevragen groepeert per soort proefpersonen. Als voorbeeld ziet men in de Tabellen 4 en 5 de antwoorden op respectievelijk de vragen 3 en 7, gegeven door de proefpersonen van panel I tijdens de beoordeling van de apparaten uit groep I.

Uit Tabel 4 blijkt dat de 'A scheerders' en de 'C scheerders' het goed eens zijn over het feit dat men zich met typen 1 en 4 gladder scheert dan met typen 2 en 3. De 'B scheerders' echter zijn van mening dat de typen 2 (een 'B apparaat'!) en 4 aanmerkelijk beter zijn dan de apparaten 1 en 3. Hun oordeel over apparaat 1 is wel zeer sterk afwijkend van dat der andere proefpersonen!

Nog sterker blijkt de duidelijke invloed van de aan het onderzoek voorafgaande scheergewoonte uit Tabel 5, waarin men ziet dat de 'A scheerders' wat betreft huidirritatie niets bijzonders merken aan type 1 (een 'A apparaat'). De 'B scheerders' en de 'C scheerders' daarentegen kwalificeren apparaat 1 als buitengewoon sterk schrijnend! Hiermede is vraag 4 bevestigend beantwoord.

Conclusies

1. Een subjectief kwaliteitsonderzoek, gebaseerd op enquêtering van proefpersonen, en uitgevoerd zoals hiervoor omschreven, kan bruikbare resultaten ten aanzien van de scheerkwaliteit van droogscheerapparaten opleveren, mits rekening wordt gehouden met het feit dat men slechts *relatieve* informatie verkrijgt, in die zin dat men alleen te weten komt hoe een bepaald type zich verhoudt tot een aantal andere apparaten waarmede het desbetreffende type 'toevallig' wordt vergeleken.

2. Uit de resultaten blijkt dat een panel van 24 personen groot genoeg is om een representatief en goed reproduceerbaar oordeel te verkrijgen. Dit neemt uiteraard niet weg dat men een onderzoek soms moet uitbreiden teneinde meer gedetailleerde conclusies te krijgen.

3. Het blijkt dat een zeer sterk 'herinneringseffekt' optreedt, hetgeen betekent dat men eenzelfde type apparaat niet tweemaal achtereen aan eenzelfde panel proefpersonen ter beoordeling mag voorleggen, in elk geval niet dan nadat er een voldoend lange 'ontwenningsperiode' in acht is genomen.

4. Zeer duidelijk blijkt dat de aan het onderzoek voorafgaande scheergewoonte van de proefpersonen een grote invloed heeft op hun oordeel. Het blijkt dat men een apparaat van het type, waarmee men zich gewend is te scheren, zonder uitzondering bijzonder goed waardeert, ongeacht de kwaliteit van de concurrentieapparaten waarmee het wordt vergeleken. Dit betekent, zoals reeds eerder werd gesteld, dat een panel proefpersonen met de grootste zorg moet worden

samengesteld, waarbij er in een 'consumentenonderzoek' beslist voor moet worden gezorgd dat elk merk even sterk in het panel is vertegenwoordigd. Op grond van de resultaten komt het ons voor, dat het in elk geval zinvol is het eindoordeel over een bepaald type te spitsen in een 'concurrentie-oordeel' en een 'bezitters-oordeel'.

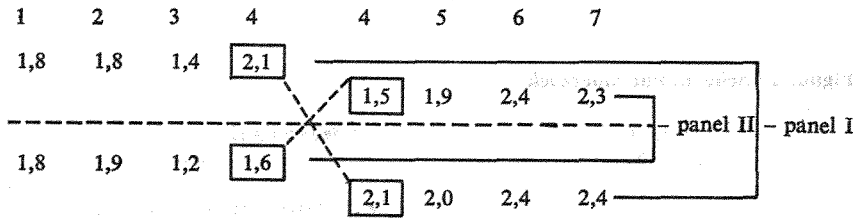
Referentie

A. J. Duncan: 'Quality control and industrial statistics', 636-639.

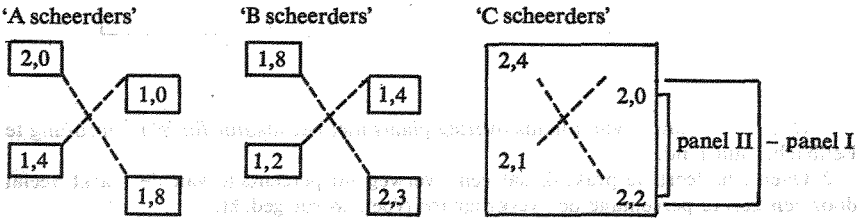
Tabel 1 / Apparatengroep I panel II periode 2

vraag	gemiddelde scores in dalende volgorde			
3	2	1	4	3
(eindgladheid)	2,0	1,9	1,8	1,2
5	2	4	1	3
(bereikbaarheid)	2,5	1,7	1,7	1,6
6	1	2	4	3
(lange haren)	1,5	1,1	1,0	0,7
7	4	3	2	1
(huidirritatie	2,1	2,0	2,0	0,7

Tabel 2 / Resultaten beoordeling scheerresultaat



Tabel 3 / Resultaten beoordeling scheerresultaat van apparaat 4



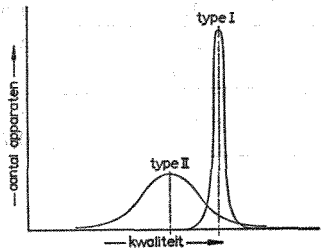
Tabel 4 / Beoordeling van de eindgladheid (vraag 3) van de apparaten uit groep I door panel I

	1(A)	2(B)	3(C)	4(C)
'A scheerders'	2,6	1,4	1,3	2,3
'B scheerders'	0,9	2,2	1,5	2,1
'C scheerders'	2,3	1,9	1,9	2,4

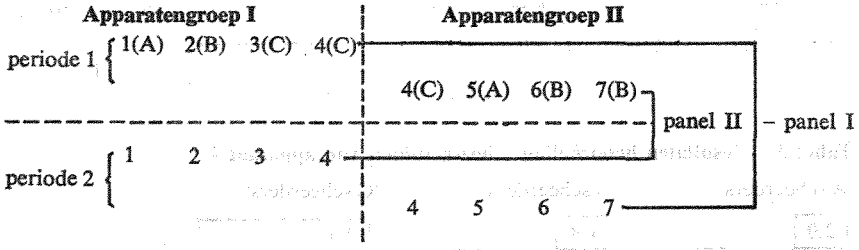
Tabel 5 / Beoordeling van de huidirritatie (vraag 7 van de apparaten uit groep I door panel I

	1(A)	2(B)	3(C)	4(C)
'A scheerders'	1,6	1,5	2,0	1,9
'B scheerders'	–1,3	2,3	1,7	2,2
'C scheerders'	–1,0	1,7	1,1	1,9

Figuur 1 / Hypothetische kwaliteitsverdelingscurve voor apparatentype I en II



Figuur 2 / Schema van onderzoek



1 Over een en ander vindt thans overleg plaats met het Institut für Warenprüfung te Fellbach, Duitsland.

2 Overigens leert de praktijk dat een overwegend percentage van de markt veelal door een gering percentage der verkrijgbare typen wordt gedekt.

De produkttest als brug tussen de laboratoriumwaarneming en het consumentenoordeel

Drs. E. N. Koster

Toepassing van de postale produkttest als techniek bij de ontwikkeling van nieuwe produkten of bij verbetering van bestaande produkten zal in het algemeen tot doel hebben om aan de hand van de uitspraak van de consument en met inachtneming van andere factoren zoals grondstoffenprijs e.d., een beslissing te vergemakkelijken of A dan wel B gekozen moet worden.

In het algemeen is dit waar in al die gevallen waarbij het te testen produkt als zodanig centraal staat.

De researchman, die in het laboratorium bezig is een produkt te ontwikkelen heeft een geheel eigen beslissingsproblematiek. Voor hem is het vaak niet voldoende om te weten dat A de voorkeur krijgt boven B zolang hij deze voorkeur niet direct kan relateren aan een of andere chemische of fysische variabele van het produkt.

De voorkeursomschrijvingen van de huisvrouwen verwarren hem eerder dan dat ze richtinggevend zijn. Wat hij behoeft is een technische 'guidance' die bij elke fase in de ontwikkeling antwoord geeft op de vraag of hij in de door hem voorgestelde richting blijft koersen.

Hij zal daarom zo mogelijk gebruik willen maken van eigen instrumentarium dat hem snel en zuiver deze 'guidance' kan verschaffen. Echter deze vorm van objectieve richtinggeving is pas zinvol als die de realiteit weerspiegelt, m.a.w. als de objectieve meetresultaten van het instrumentarium in relatie te brengen zijn met de subjectieve opinies van de consumenten.

Zo worden bij de ontwikkeling van waspoeders witheidseffekten gemeten met behulp van lichtreflectie.

Het verschil in de hoeveelheid gereflecteerd licht voor en na de was representeert het verschil in witheid. De verhouding van dit verschil en de maximaal mogelijke toename in witheid geeft de wasefficiency en zal ergens liggen tussen de 0 en 100%.

De wijze waarop in het laboratorium het stukje wasgoed wordt gewassen,

waarop de meting wordt verricht, behoeft natuurlijk niet overeen te komen met de methode die door de gemiddelde huisvrouw wordt toegepast.

De vraag die de researchers van het laboratorium dan ook zullen stellen is of er enige relatie is tussen hun objectieve metingen en de subjectieve meting van de huisvrouw. Bovendien kan de vraag gesteld worden hoe groot het verschil in wasefficiency moet zijn om door de huisvrouw opgemerkt te worden.

Op deze vragen kon een antwoord worden gegeven door gebruik te maken van de methode van de postale produkttest.

De testopzet was als volgt. Huisvrouwen van een consumentenpanel kregen twee samenstellingen waspoeder toegestuurd welke zij successievelijk op verschillende wasdagen moesten gebruiken voor haar normale huishoudwas. Het essentiële verschil met een normale produkttest was hierin gelegen dat de huisvrouw een proefdoekje moest meewassen. Het proefdoekje was gedeeltelijk geïmpregneerd met een oplossing van vettige bestanddelen en gedeeltelijk met een thee-oplossing. Daar de test een paarsgewijze vergelijking inhield, kreeg de huisvrouw 2 proefdoekjes, voor elk testprodukt één.

Om de test niet te gecompliceerd te maken waren de waspoeders zodanig samengesteld dat ze slechts t.a.v. 2 componenten qua concentratie van elkaar verschilden. Voor de ene variabele waren 2 concentratieniveaux gekozen, voor de andere variabele 3 niveaux.

De proefsamenstellingen werden in een reeks van paarsgewijze combinaties getest zodat uit de resultaten de invloed op de preferentie van de niveaux van beide variabelen kon worden vastgesteld.

Elke huisvrouw, behorende tot een van de onderling vergelijkbare steekproeven, kreeg slechts één combinatie van 2 proefsamenstellingen te keuren.

De met de testprodukten toegezonden vragenlijst zag er betrekkelijk eenvoudig uit en centreerde zich om de preferentievraag die 2 maal werd gesteld. De huisvrouw moest nl. de beide produkten beoordelen voor wat betreft het effect op de meegewassen proefdoeken en wel één maal t.a.v. het gedeelte dat met vet bevuild was en één maal t.a.v. het theegedeelte.

De doeken moest zij terugzenden. Deze werden vervolgens door het laboratorium doorgemeten voor de vaststelling van de absolute wasefficiency (= W.E.).

Voornaamste resultaten

Voor de analyse werd voornamelijk gebruik gemaakt van 2 gegevens, enerzijds de absolute W.E. (door het laboratorium vastgesteld) gebaseerd op de wasresultaten van de proefdoeken en anderzijds de subjectieve voorkeur uitgesproken door de huisvrouw.

Deze 2 resultaten konden per huisvrouw in dezelfde richting gaan maar de

relatie kon ook tegengesteld zijn, m.a.w. de huisvrouw kon bijv. A prefereren terwijl het absolute verschil in wasresultaat A-B nul of negatief was.

De onderstaande tabel laat de subjectieve voorkeur zien gekruistabuleerd naar de verschillen in wasefficiency van beide proefdoekjes welke de huisvrouw moest gebruiken. De produkt-identificatie is hier van ondergeschikt belang omdat het in dit stadium uitsluitend gaat om het mogelijk vaststellen van algemene tendenzen.

Tabel 1 / Verhouding subjectief oordeel van de huisvrouw en de objectieve meting van het wasresultaat

		Absolute verschil in W.E. tussen 2 produkten							
	Totaal	Geen verschil	0.1-4.9	5.0-9.9	10.0-14.9	15.0-19.9	20.0-24.9	25.0-29.9	30.0 en hoger
Subjectieve voorkeur									
Zeer veel beter	12%	(1)	5%	8%	18%	25%	31%	50%	(2)
Beslist beter	28%	–	15%	29%	37%	51%	53%	38%	(5)
Alleen iets beter	31%	(1)	35%	37%	26%	21%	12%	12%	(1)
Alle respondenten met voorkeur voor produkt met hoger W.E	70%	(2)	54%	74%	81%	96%	97%	100%	(8)
Zeer veel beter	1%	–	1%	–	–	–	–	–	–
Beslist beter	4%	–	5%	4%	3%	2%	–	–	–
Alleen iets beter	10%	–	18%	5%	5%	2%	3%	–	–
Alle respondenten met voorkeur voor produkt met lagere W.E.	14%	–	25%	10%	8%	4%	3%	–	–
Geen voorkeur	15%	(3)	21%	16%	11%	–	–	–	–
Alle ondervraagden	(740)	(5)	(299)	(221)	(98)	(53)	(32)	(24)	(8)

Terwille van de eenvoud zijn hierin alleen de cijfers opgenomen die betrekking hebben op het 'vet'-gedeelte van de proefdoeken. Overigens liepen de resultaten van het thee-gedeelte hieraan parallel.

Commentaar

70% van de respondenten had een voorkeur die in overeenstemming was met het positieve verschil in W.E. Met andere woorden, de huisvrouw had de juiste keuze gemaakt; 14% bleek een voorkeur te hebben voor het produkt dat een negatief verschil in W.E. te zien gaf, dus een lagere absolute wasefficiency had; 15% had geen voorkeur ondanks het feit dat er wel degelijk verschil in absoluut wasresultaat was gemeten.

Uit de resultaten zijn de volgende conclusies te trekken:

Huisvrouwen blijken in staat om te discrimineren; een verschil van 5.0-9.9 eenheden in W.E. blijkt voldoende om het percentage juiste antwoorden aanzienlijk te doen stijgen. Zelfs kan gesteld worden dat het verschil van ca. 5 eenheden min of meer de kritieke waarde vertegenwoordigt, waarbeneden het moeilijk is voor de huisvrouw om verschil te ontdekken.

Een 2e conclusie is dat een bewijs is gevonden voor de intrinsieke betekenis van het verschil in voorkeursintensiteit uitgedrukt in de bewoordingen 'zeer veel beter', 'beslist beter' en 'alleen iets beter'. Uit voorstaande tabel is gemakkelijk een correlatie te ontdekken tussen de voorkeursintensiteit en het absolute verschil in W.E. Hoe hoger het verschil in W.E., des te sterker de voorkeursintensiteit. Deze conclusie betekent uiteraard niet dat het bestaan ontkend wordt van psychologische verschillen tussen huisvrouwen welke van invloed zijn op de wijze waarop zij zich uitdrukken (de ene huisvrouw zal eerder 'zeer veel beter' aankruisen dan de andere op grond van eigen instelling), echter mag gesteld worden dat een relatie gevonden is tussen objectieve bevindingen en subjectieve meningen.

Om vervolgens aan te tonen dat er enige positieve relatie bestaat tussen de bevindingen op het laboratorium en de huisvrouwelijke voorkeur – uitgesproken onder normale produkttest-omstandigheden waarbij dus geen proefdoek als meetlat wordt gebruikt – werden de resultaten van deze tests vergeleken met die afkomstig uit eerder gehouden tests met volledig vergelijkbare testprodukten.

Het bleek dat de niveauverschillen van de geteste produktvariabelen ook onder normale produkttest-omstandigheden konden worden waargenomen, echter met mindere precisie.

De fractie huisvrouwen met een voorkeur bleek bepaald kleiner in de normale testsituatie, ruwweg in de orde van 60-70% tegen 70-90% in de tests met de proefdoeken.

Ook werd een lager netto voorkeurspercentage gevonden (verschil in voorkeur voor A en B); ca. 9-20% in normale testomstandigheden tegen 20-40% in de tests met proefdoeken.

Samenvattend mag gesteld worden dat het mogelijk is gebleken om met behulp van de techniek van de produkttest het laboratorium informatie te verschaffen omtrent de relatie tussen subjectieve voorkeur en objectieve meetresultaten. Doordat het laboratorium aan de lichtreflectie-meetmethode predictieve waarde mag toekennen wordt bereikt dat niet in elke fase van de produktontwikkeling de subjectieve mening van de huisvrouw behoeft te worden gepeild.

Het indiceren van de 'drempelwaarde' maakt het mogelijk om het overbodig testen van produkten, waarvan de verschillen toch niet door huisvrouwen kunnen worden waargenomen, uit te bannen.

Een van de basis-problemen bij produkt-testing: de 'Friendliness-Bias'

Kritische aantekeningen bij de lezing van prof. Van Tulder tijdens het ESOMAR Congres 1969

W. van Vliet

Inleiding

Bij een produkt-test wordt gepoogd de acceptatie van een test-produkt te meten. Hierbij wordt gewoonlijk het oordeel van een groep van potentiële consumenten gevraagd over het test-produkt en één of meer referentie-produkten. Een referentie-produkt is dan een produkt waarvan de acceptatie bekend is, dus bijv. een bestaand produkt met een behoorlijk marktaandeel. Het gaat er niet in de eerste plaats om voor het testprodukt een absoluut waarde-oordeel te verkrijgen, maar veeleer een relatief beeld. Teneinde een onvertekend relatief beeld te verkrijgen, dient aan twee voorwaarden te worden voldaan:

a. De proefpersoon dient in staat te worden gesteld een onpartijdig oordeel te vellen, d.w.z. dat bij aanvang der test de proefpersoon onbevooroordeeld dient te staan tegenover test-produkt en referentie-produkt(en).

b. Test-produkt en referentie-produkt(en) dienen op gelijke wijze en onder gelijke omstandigheden te worden geconsumeerd.

Aan de eerste voorwaarde is moeilijk geheel te voldoen. Een proefpersoon wie gratis een artikel ter consumptie wordt aangeboden zal vaak geneigd zijn hierover een positief oordeel te geven, omdat het nogal ondankbaar lijkt een zeer negatief geluid te laten horen over iets dat men ten geschenke heeft gekregen, vooral als dat iets een vrij grote geldswaarde vertegenwoordigt.

Deze neiging om een te positief oordeel te geven is de zgn. 'Friendliness-bias'. Iedere marktonderzoeker die met produkt-tests te maken heeft weet dat deze friendliness-bias een belangrijke hindernis kan vormen op de weg naar de juiste interpretatie van de verkregen test-resultaten.

Voor een bestaand produkt geldt voorts dat het oordeel van de consument niet alleen wordt bepaald door de produkt-eigenschappen (smaak, reinigingskracht, etc.), maar ook door merknaam, fabrikant, verpakking, hoeveelheid, prijs, etc. Wenst men dus de produkteigenschappen van een onbekend nieuw merk te testen en gebruikt men als referentie-produkt een befaamd, bestaand merk in de originele verpakking, dan mag niet zonder meer verwacht worden dat de proef-

personen in gelijke mate onbevooroordeeld tegenover test-produkt en referentie-produkt staan.

Als tweede voorwaarde voor het verkrijgen van een onvertekend relatief beeld werd gesteld, dat test-produkt en referentie-produkt(en) op gelijke wijze en onder gelijke omstandigheden dienen te worden geconsumeerd. Aan deze voorwaarde wordt in de praktijk vaak te weinig aandacht besteed.

Indien bijv. op straat of in een warenhuis een smaaktest wordt gehouden voor een nieuw produkt en het relatieve oordeel van de consument hierbij wordt gevraagd, dan wordt meestal verwaarloosd dat het tijdstip waarop het test-produkt wordt geconsumeerd in belangrijke mate kan afwijken van de tijdstippen waarop het referentie-produkt gewoonlijk wordt geconsumeerd.

De meervoudige produkt-test

Een veelvuldig toegepaste methode van produkt-testing is die waarbij de proefpersoon zowel het test-produkt als één of meer referentie-produkten krijgt aangeboden. Indien de produkten in gelijksoortige (bij voorkeur blinde) verpakkingen worden aangeboden dan mag worden aangenomen dat vooroordelen geen of weinig invloed zullen hebben op de te maken keuze.

Bij deze methoden wordt de friendliness-bias als zodanig niet uitgeschakeld, doch wordt aangenomen dat deze friendliness-bias in ongeveer gelijke mate het oordeel over test-produkt en referentie-produkt(en) beïnvloedt, zodat bij onderlinge vergelijking der resultaten voor test-produkt en referentie-produkten de invloed van de friendliness-bias geacht kan worden min of meer te zijn geëlimineerd. Een voordeel van deze testmethode is voorts, dat er voor kan worden gezorgd dat test-produkt en referentie-produkten (vrijwel) gelijktijdig en onder gelijke omstandigheden worden geconsumeerd, zodat ook in dit opzicht belangrijke vertekening van de uitkomsten kan worden voorkomen.

Voor de goede orde dient hier te worden aangetekend, dat de verkregen uitkomsten geen *absolute* waarde hebben. Indien bijv. 40% van de proefpersonen te kennen geeft bereid te zijn het test-produkt te kopen dan kan met behulp van dit cijfer *geen* prognose voor het aantal toekomstige consumenten worden gemaakt, doch wel worden geconstateerd of de koopbereidheid kleiner, gelijk of groter is dan van een referentie-produkt.

De enkelvoudige produkt-test

Indien een proefpersoon alleen het test-produkt ter consumptie krijgt aangeboden geldt als referentie-produkt veelal het merk dat gewoonlijk wordt gebruikt of het

merk dat het laatst werd gebruikt. In verband met het voorgaande is het duidelijk dat aan deze testmethode belangrijke nadelen zijn verbonden nl.:

1. De friendliness-bias geldt nu uitsluitend het test-produkt waardoor de verkregen vergelijking tussen test-produkt en referentie-produkt misleidend kan zijn.
2. Het gevaar bestaat dat het test-produkt niet op gelijke wijze en niet onder gelijke omstandigheden wordt geconsumeerd als het referentie-produkt.

Men kan zich afvragen waarom, gezien de nadelen, toch nog zo vaak gebruik gemaakt wordt van enkelvoudige produkt-tests. Voor bepaalde produkten is hiervoor de verklaring eenvoudig nl. dat een meervoudige test niet of nauwelijks mogelijk is. Stel bijv. dat men de acceptatie van een nieuw chocolade-produkt wil meten en dat dit nieuwe produkt nauwelijks lijkt op één van de bestaande chocolade-produkten. In dat geval is een zinnige keuze van een referentie-produkt niet eenvoudig. Daarom wordt een enkelvoudige produkt-test ingesteld en als referentie-produkten worden dan de chocolade-produkten genomen die de proefpersoon wel eens eet (welke produkten dit zijn wordt uiteraard vooraf voor elke proefpersoon afzonderlijk vastgesteld).

De resultaten die voor het test-produkt op deze wijze worden verkregen kunnen niet zonder meer worden geïnterpreteerd. Indien echter een serie van soortgelijke onderzoekingen wordt doorgevoerd, steeds met dezelfde referentie-produkten doch met verschillende test-produkten (waaronder ook één of meer bestaande produkten), dan worden de uitkomsten voor de test-produkten uit verschillende onderzoekingen wel onderling vergelijkbaar.

Hierbij dient te worden opgemerkt, dat de vergelijking tussen de resultaten voor twee test-produkten uit twee verschillende onderzoekingen statistisch minder nauwkeurig is dan ingeval de twee produkten deel uitmaken van één onderzoek.

Er zijn dus gevallen dat een enkelvoudige produkt-test de enige oplossing is. Er wordt echter ook vaak een enkelvoudige produkt-test ten onrechte uitgevoerd, nl. in die gevallen waar een meervoudige test mogelijk zou zijn en veel betrouwbaarder resultaten zou opleveren.

Gevreesd moet worden dat de oorzaak hiervan gezocht moet worden in onvoldoende inzicht in de problematiek van de produkt-test, dat bij een deel der opdrachtgevers en bij bepaalde bureaux aanwezig schijnt te zijn. Dit gebrek aan inzicht werd onlangs treffend geïllustreerd door de lezing die Prof. Van Tulder tijdens het in Amsterdam gehouden ESOMAR Congres ten beste gaf. In deze lezing werd een experimenteel onderzoek beschreven, dat door de N.V. v/h Nederlandse Stichting voor Statistiek werd uitgevoerd. Het doel van dit experiment was om voor een enkelvoudige produkt-test:

a. de grootte van de friendliness-bias vast te stellen,
b. een methode te vinden om deze friendliness-bias aanzienlijk te reduceren,
c. een praktische methode te vinden om de invloed van de friendliness-bias op de resultaten geheel te elimineren.

Volgens het betoog van Prof. Van Tulder heeft het experiment tot de beoogde resultaten geleid.

Gezien het belang van dit onderwerp zullen hieronder een aantal punten van dit experiment nader onder de loupe worden genomen.

Opzet van het experiment

De steekproef bestond uit huisvrouwen, die het weekmiddel Biotex groen gebruiken en die de was geheel of *gedeeltelijk* thuis doen. In de basistest (A) kregen de huisvrouwen een normale verpakking Biotex groen aangeboden met de mededeling dat de inhoud van de verpakking niet hetzelfde is als wat ze in gebruik heeft. Deze huisvrouwen kregen een week de tijd om het test-produkt te proberen. In werkelijkheid kregen de huisvrouwen het normale produkt Biotex groen. Prof. Van Tulder stelt nu '*Alle pakken die in de test werden gebruikt bevatten normaal Biotex groen, zodat de testresultaten niet aan produktverschillen kunnen worden toegeschreven*'. Dit op zichzelf is uiteraard juist, maar dat derhalve een eventuele voorkeur voor het test-produkt geheel aan de friendliness-bias dient te worden toegerekend is niet juist.

De huisvrouwen werd nl. gesuggereerd dat het ging om een vernieuwd en (vermoedelijk) verbeterde soort Biotex. Aan de in het begin van dit artikel gestelde voorwaarde van onbevooroordeeldheid bij de aanvang der test is dus allerminst voldaan. Deze suggestie werd nog aanzienlijk versterkt door de eerste testvragen. Onmiddellijk na de uitreiking werd nl. een visuele vergelijking tussen test-produkt en de in huis aanwezige Biotex gemaakt. Ten aanzien van de *geur* leidde dit tot de volgende uitkomsten:

voorkeur voor test-produkt	71%
voorkeur voor eigen produkt	10%
geen voorkeur	19%
n =	(87)

Deze overweldigende voorkeur voor het test-produkt doet Prof. Van Tulder toegeven dat 'aangenomen mag worden dat de geur van een pas geopend pak altijd beter is dan de geur van een pak dat al enige tijd open is'.

Door de gebruikte vraagstelling wordt voor de huisvrouw als het ware bewezen dat het nieuwe produkt beter is dan het oude. Nochtans gaat Prof. Van Tulder rustig door met te beweren dat de test-resultaten een weerspiegeling zijn van de friendliness-bias.

Bij de eerste confrontatie test-produkt – eigen produkt werd behalve over de geur een oordeel gevraagd over 'algemene indruk', 'korreligheid' en 'kleur'. Ook hier kreeg het test-produkt aanzienlijk meer stemmen dan het eigen produkt; de

vraag is echter weer of het uiterlijk van een pasgeopend pak niet in allerlei opzichten aantrekkelijker is dan het uiterlijk van een gebruikt pak. Het behoeft geen betoog dat na deze confrontatie de produkteigenschappen van het testprodukt na gebruik beter beoordeeld werden dan die van het normaal gebruikte produkt. Als er een prijsvraag zou worden uitgeschreven 'Hoe krijg ik de grootste vertekening in produkt-test resultaten', dan zou dit experiment een kanshebbende inzending zijn. Het sukses zou misschien nog vollediger zijn geweest indien de huisvrouw er bij was verteld dat het test-produkt tweemaal zo duur is als het normale produkt.

Met een meting van de friendliness-bias heeft een en ander echter niets te maken. Eén nuttige en belangrijke lering kan echter uit dit deel van het experiment worden getrokken: ogenschijnlijk geringe fouten in de opzet kunnen bij een produkt-test grote vertekeningen in de test-resultaten teweegbrengen.

Het reduceren van de Friendliness-bias

Prof. Van Tulder zegt een methode gevonden te hebben om de bij test A gevonden bias ongeveer te halveren. Hiertoe werd bij het experiment een drietal groepen huisvrouwen ondervraagd als bij groep A, evenwel werd bij de uitreiking van het test-produkt als volgt afgeweken:

Groep B: de huisvrouw kreeg de keus uit 3 gelijke pakken, één met het 'oude' produkt en twee met het 'nieuwe' produkt.

Groep C: de huisvrouw kreeg de keus uit 2 gelijke pakken, waarvan één met het oude en één met het nieuwe produkt.

Groep D: als B maar nu twee oude en één nieuwe.

In werkelijkheid werd in alle gevallen het 'oude' produkt gegeven. In alle 3 groepen is de voorkeur voor het test-produkt geringer dan bij groep A; het beste resultaat wordt geboekt bij groep B.

Vergelijking tussen groep A en groep B voor wat betreft de resultaten van de confrontatie van het test-produkt met het normaal gebruikte produkt (met uitzondering van de geur) geeft het volgende beeld:

Tabel 1	Algemene indruk		Korreligheid		Kleur	
	Groep A %	Groep B %	Groep A %	Groep B %	Groep A %	Groep B %
Voorkeur voor test-produkt	30	18	35	25	35	24
Voorkeur voor eigen produkt	4	6	14	16	10	9
Geen voorkeur	66	76	51	59	55	67
n =	(87)	(88)	(87)	(88)	(87)	(88)

Prof. Van Tulder stelt: 'indien er geen friendliness-bias zou zijn zou het verschil tussen het eerste en tweede cijfer ongeveer nul zijn en het derde cijfer (geen voorkeur) ongeveer 100%'.

Het eerste deel van deze stelling is juist, het tweede deel echter niet. Dat een bepaald deel der ondervraagden een verschil meent te zien waar geen verschil is heeft niets te maken met de friendliness-bias maar weerspiegelt het beperkte onderscheidingsvermogen van de mens. Ook zonder friendliness-bias gokt een deel der ondervraagden dat er wel enig verschil zal zijn. Trouwens in het onderhavige geval was er waarschijnlijk in een deel der gevallen wel degelijk een onderscheid te zien. Dat bij groep B het percentage dat geen voorkeur heeft steeds hoger is dan bij groep A is zeker niet geheel toevallig; immers bij groep A werd gezegd dat het produkt anders is, bij groep B werd gezegd dat het testprodukt mogelijk het eigen produkt is. Hierdoor hadden de huisvrouwen die deel uitmaakten van groep B de mogelijkheid te constateren dat ze het eigen produkt hadden ontvangen, voor groep A is deze mogelijkheid niet aanwezig.

Prof. Van Tulder definieert de friendliness-bias als het percentage dat het testprodukt prefereert, verminderd met het percentage dat het eigen produkt prefereert. Op deze wijze worden uit tabel 1 de volgende percentages afgeleid:

Tabel 2	**Groep A**	**Groep B**
Algemene indruk	30 − 4 = 26	18 − 6 = 12
Korreligheid	35 − 14 = 21	25 − 16 = 6
Kleur	35 − 10 = 25	24 − 9 = 15

De uitkomsten voor groep B zijn ongeveer de helft van die voor groep A dus komt Prof. Van Tulder tot de slotsom dat in groep B de friendliness-bias tot de helft van die van groep A is verminderd. Deze conclusie is echter aanvechtbaar. Indien er geen friendliness-bias zou zijn zouden de beide produkten ongeveer een gelijk aantal voorkeurstemmen hebben verkregen.

Het gaat dus om de *verhouding* waarin de voorkeurstemmen over de beide produkten zijn verdeeld; degenen die geen voorkeur uitspreken dienen buiten beschouwing gelaten te worden.

Zo kan met behulp van tabel 1 de volgende opstelling worden gemaakt:

Tabel 3	**Algemene indruk**		**Korreligheid**		**Kleur**	
	Groep A %	**Groep B** %	**Groep A** %	**Groep B** %	**Groep A** %	**Groep B** %
Voorkeur voor testprodukt	88	76	71	61	77	72
Voorkeur voor eigen produkt	12	24	29	39	23	28
Steekproefgrootte *	26	21	38	36	35	29

* Degenen die geen voorkeur uitspraken zijn hier steeds buiten beschouwing gelaten.

Hoewel ook hier de cijfers voor het testprodukt voor groep B lager zijn dan voor groep A zijn de verschillen statistisch op geen stukken na significant.

Praktische toepassing

In zijn conclusie komt Prof. Van Tulder onder meer tot de volgende suggestie:

'bij het testen van een nieuw of vernieuwd produkt laat men de proefpersoon uit 3 verpakkingen kiezen na gezegd te hebben dat één van de 3 verpakkingen het eigen produkt bevat; op deze wijze wordt nl. de friendliness-bias tot een minimum teruggebracht'.

Hoe dit in de praktijk werkt is niet duidelijk; immers het is uiterst onwaarschijnlijk dat het test-produkt er ongeveer hetzelfde uitziet als het eigen produkt, zodat de proefpersoon gewoonlijk onmiddellijk kan zien dat het een ander produkt is. Dit betekent dan dat de gedane suggestie zinloos is en het systeem niet *kan* werken.

Tenslotte stelt Prof. Van Tulder dat door vergelijking van de verkregen uitkomsten voor het nieuwe produkt met die uit een controlesteekproef (waarbij volgens systeem B van het experiment wordt gewerkt) de opgetreden friendliness-bias kan worden geëlimineerd. Ook al zou het systeem in de praktijk kunnen worden doorgevoerd en ook al zou de bij het experiment gevonden vertekening werkelijk aan friendliness-bias zijn toe te schrijven dan nog moet achter deze bewering een groot vraagteken worden geplaatst. Dit kan wellicht het best met een voorbeeld worden verduidelijkt. Stel dat Biotex een nieuw soort waspoeder wil testen; ongetwijfeld zal dan het percentage 'geen voorkeur' veel lager uitvallen dan in de controlegroep (groep B van het experiment). Stel dat bijv. voor kleur de volgende resultaten worden verkregen:

	Eigenlijke test	**Controletest**
	%	%
voorkeur voor test-produkt	60	24
voorkeur voor eigen produkt	30	9
geen voorkeur	10	67

Gaat men te werk volgens het systeem van Prof. Van Tulder dan krijgt men:

voorkeur testprodukt boven eigen produkt – eigenlijke test 30%
controletest 15%

Ergo, de helft van het verschil van 30 is slechts te verklaren uit de friendliness-bias dus het nieuwe test-produkt wordt werkelijk geprefereerd.

Kijkt men echter naar de verhouding waarin beide produkten worden geprefereerd dan is deze verhouding voor het test-produkt ongunstiger in de eigen-

lijke test dan in de controletest, zodat geconcludeerd zou moeten worden dat het eigen produkt wordt geprefereerd.

Conclusie

Het experiment zoals dat is uitgevoerd voldoet qua opzet, qua wijze van interpreteren van de resultaten en qua gemaakte gevolgtrekkingen, in de verste verte niet aan de eisen die men aan een verantwoord experiment mag stellen.

Antwoord op de kritische aantekeningen van de heer Van Vliet

Prof. dr. J. J. M. van Tulder

Bij het totstandkomen van een resultaat van een producttest spelen vele variabelen een rol. Tot deze vele variabelen behoren er ook die weinig of niets met het product zelf te maken hebben. Een dergelijke variabele is de wijze waarop het testproduct wordt aangeboden. Hoe ondergaat de proefpersoon de aanbieding en wat is daarvan de invloed op het resultaat van de test?

Het is bekend, dat er ten aanzien van het aangeboden product zekere verwachtingen bestaan bij de proefpersoon. Hij of zij moet wel veronderstellen, ook al wordt dit bij het aanbieden niet expliciet gezegd, dat het testproduct op een of meer punten afwijkt van het product waarmede het wordt vergeleken. Het door mij opgezette experiment had *alleen* ten doel, na te gaan wat er gebeurt als de verwachtingen van de proefpersoon op een bepaalde wijze worden beïnvloed tijdens de aanbieding. Voor het experiment werd dus alleen dit aspect gevarieerd. Complicaties door het inbrengen van andere aspecten werden zoveel mogelijk vermeden.

Het probleem van de enkelvoudige of de meervoudige test kwam dan ook niet aan de orde. Bij dit experiment werd gekozen voor het enkelvoudige systeem. Bij een mogelijk volgend experiment kan het meervoudige systeem worden toegepast, alhoewel dit de testopzet wel enigszins compliceert, zeker bij een product als een voorwasmiddel.

De heer Van Vliet zegt in zijn kritiek, dat, volgens mij, het doel van het experiment drieledig was. Ik meen dat hier sprake moet zijn van een verkeerde interpretatie van mijn tekst, aangezien het zeker niet direct de bedoeling is geweest om het derde doel, namelijk met dit ene experiment een praktische methode te vinden om de invloed van het vriendelijkheidseffect geheel te elimineren, te bereiken. Het experiment is een *stap* op de weg om een reductie van het effect te verkrijgen en misschien is het einde van de weg wel de eliminatie. Zover zijn wij echter nog niet.

Nu het experiment zelf. In de groepen B, C en D werd door de enquêteur/trice

gezegd dat het eigen product of een ander product in de pakken kon zitten, waarbij de kans om het eigen product als testproduct te krijgen voor de drie groepen verschillend was. Er moest echter over een *ander* product worden gesproken. In groep A is dit, om een goede vergelijking te krijgen ook gebeurd. Dit kan bovendien geen kwaad, want – zoals reeds gesteld – een proefpersoon heeft altijd de verwachting dat hij of zij een ander product krijgt aangeboden. Geen ondervraagde zal – ook als er niets wordt gezegd – ooit aannemen, dat bij het aanbieden van een testproduct, dit product volkomen identiek zal zijn met hetgeen hij of zij reeds gebruikt. Dit is te ongeloofwaardig. Er werd bij het aanbieden niets gezegd dat er op kon duiden dat het testproduct *beter* zou zijn. Juist dit punt is belangrijk, want het vriendelijkheidseffect resulteert in een systematisch 'beter' vinden van het testproduct. De geur-resultaten, die de heer Van Vliet vervolgens bespreekt, zijn ook voor mij niet erg duidelijk. De geur is meestal een zeer moeilijk punt bij producttests. Ik heb verondersteld dat de versheid van de geur bij het openen van het pak mede een rol speelt. Met de heer Van Vliet ben ik echter ook van mening dat de uitkomst vreemd blijft. De fabrikant heeft op mijn verzoek een speciale controle uitgeoefend wat betreft de geur. Daarbij is mij verzekerd, dat er absoluut geen verschil was tussen de kwantiteit en de kwaliteit van het gebruikte parfum voor het gewone product en het testproduct. Wat betreft andere aspecten dan de geur ben ik het niet eens met de heer Van Vliet. Dat bijv. korreligheid en kleur bij een pas geopend pak voorwasmiddel aantrekkelijker zijn dan bij een gebruikt pak komt mij zeer onwaarschijnlijk voor, zeker indien wij bedenken dat dit soort middelen tamelijk snel wordt verbruikt en ouderdom geen grote rol zal spelen.

Wat betreft de nuttige en belangrijke lering die de heer Van Vliet uit het experiment wil trekken geloof ik wel dat hij gelijk heeft. Echter hiervoor was geen experiment nodig.

Indien er geen vriendelijkheidseffect zou zijn en men zou met proefpersonen te maken hebben die *kunnen* onderscheiden – hetgeen zelden het geval is – dan zou de verdeling theoretisch als volgt zijn:

voorkeur voor testprodukt	0%
geen voorkeur	100%
voorkeur voor eigen produkt	0%
totaal	100%

Uiteraard is er wel spreiding bij een groep proefpersonen. Zonder bias moet de verdeling echter symmetrisch blijven.

Wat betreft het verschil tussen de percentages 'geen voorkeur' zou ik willen opmerken dat dit nu juist één van de resultaten van het experiment is.

Een toeneming van de 'neutraliteit' van de proefpersonen moest *bij dit experiment* tot uitdrukking komen in een stijging van het percentage 'geen voorkeur'.

Bij de significantietoets moet deze antwoordgroep dan ook worden meegenomen. Ze mag zeker niet, zoals de heer Van Vliet doet, buiten beschouwing worden gelaten.

En nu de practische toepassing. Het hypothetische voorbeeld dat de heer Van Vliet geeft lijkt mij voor discussie vatbaar. Blijft echter, dat het experiment heeft geleerd, dat het in principe mogelijk is een proefpersoon tot op zekere hoogte neutraal te stellen. Volledig lukt dit nog niet, maar nieuwe research zal ons misschien weer een stap verder brengen.

Uit het experiment blijkt dat bij een enkelvoudige producttest, indien het volgens de fabrikant om een mogelijke verbetering van een bestaand product gaat, het vriendelijkheidseffect gereduceerd kan worden.

Tenslotte de conclusie van de heer Van Vliet. Uit het bovenstaande is gebleken dat ook t.a.v. kritiek op een experiment, weer kritiek kan worden gegeven. Ik meen dan ook dat de conclusie, zoals de heer Van Vliet die trekt, qua interpretatie en gevolgtrekking in de verste verte niet verantwoord is.

Naar ik hoop zullen onze beide beschouwingen er echter toe bijdragen dat er middelen worden gevonden om door middel van samenwerking tussen producenten en onderzoekbureaus de research voor producttests in Nederland verder te ontwikkelen.

Commentaar

Hoe groot is de markt voor een nieuw produkt?

drs. A. W. Jongert

Gezien vanuit de marketing-filosofie kan men zich afvragen of het verantwoord is om belangrijke bedragen te investeren in de ontwikkeling van nieuwe producten waarvan de potentiële markt niet of onvoldoende bekend is.

Als hierop een ontkennend antwoord wordt gegeven dan komt onmiddellijk de vraag of het mogelijk is om te bepalen hoe groot de afzetmogelijkheden zijn van een product dat het publiek niet kent.

Toen in 1963 door het Dr Neher laboratorium het Centraal Antenne Systeem was ontwikkeld en er besloten was in het Haagse Bezuidenhout een proefgebied in te richten om de technische kwaliteiten van het CAS te testen, leek dit een goede gelegenheid om na te gaan in hoeverre marktonderzoek een bijdrage kan leveren aan het opstellen van een verkoopprognose voor dit nog onbekende product.

In het Bezuidenhout werd een willekeurige steekproef getrokken van 200 gezinnen.

Voordat deze op de hoogte waren gesteld van het feit dat zij binnenkort via een CAS-abonnement in staat zouden worden gesteld om zonder eigen antennes twee buitenlandse TV-stations en 15 fm-radioprogramma's op hun TV- en radiotoestel te ontvangen, werden in deze gezinnen vraaggesprekken over het CAS gehouden door enquêteurs van het NIPO.

Hierin werden alle aspecten van het CAS systematisch aan de orde gesteld waardoor de ondervraagden een duidelijk beeld van de nieuwe dienstverlening kregen.

Uit de enquête-resultaten bleek dat de levering van buitenlandse tv-programma's als veel attractiever werd beoordeeld dan de mogelijkheid van ontvangst van fm-radioprogramma's.

De prognose werd daarom gericht op de bezitters van een TV-toestel. Binnen deze groep bleek 80% geïnteresseerd in buitenlandse tv-programma's waarvan 57% voor Duitsland, 33% voor België, 18% voor Frankrijk en 17% voor Enge-

land. De aanbieding van één Duits en één Belgisch programma naast het Nederlandse kwam goed overeen met de uitgesproken voorkeur.

Door te vragen wat men over zou hebben voor aansluitkosten en abonnementsgeld is getracht na te gaan hoe groot de prijsgevoeligheid zou zijn.

Gegevens die uit deze vraagstelling werden verkregen gaven aanleiding te veronderstellen dat de prijsstelling een niet onbelangrijke invloed op de te verwachten vraag zou hebben.

Bij deze vraag gaven 119 ondervraagden een aansluitprijs op en 133 een abonnementsprijs. De onderstaande tabel geeft een indruk van de prijselasticiteit.

aansluitkosten	**perc. ondervr.** %	**abonnementskosten per maand**	**perc. ondervr.** %
		minder dan	
f 5,—	100	f 2,—	100
f 15,—	85	f 2,—	96
f 25,—	65	f 3,—	90
f 35,—	26	f 4,—	84
f 45,—	23	f 5,—	76
f 55,—	14	f 6,—	43
f 65,—	10	f 7,—	33
f 125,—	5	f 10,—	26
		meer dan	
		f 10,—	6

Om de belangstelling bij alle tv-bezitters te meten werd gesteld dat een abonnement op het CAS zou kunnen worden verkregen tegen f 20,– aansluitkosten en f 5,– abonnementskosten per maand.

Aan de ondervraagden werd een koopplanschaal voorgelegd en hun werd verzocht zich zelf hierop in te delen.

De belangrijkste uitkomsten waren:

meningen over CAS-abonnement	**alle tv-bezitters** %		**tv-bezitters met belangst. voor buitenl. zenders** %	
zeker abonneren	26	54	28	62
zeer waarschijnlijk	28		34	
weet niet, misschien	25		20	
denk het niet	15	21	13	18
zeker niet	6		5	

Op basis van deze meting kon worden geconcludeerd dat ca. 50% van alle tv-bezitters in de markt zou zijn voor een abonnement.

Begin oktober kregen alle bewoners in het proefgebied een circulaire van de PTT waarin mededeling werd gedaan over de mogelijkheid zich op het CAS te abonneren en werd alle belangstellenden verzocht zich bij de PTT op te geven.

In deze aanbieding is de aansluitprijs gesteld op f 15,– en de abonnementsprijs op f 4,– per maand.

In januari 1964 is nagegaan welke ondervraagden uit de septemberenquête reeds waren aangesloten op het CAS. Hierbij bestond de mogelijkheid dat enkele ondervraagden wel een aanvraag hadden ingediend doch dat de aansluiting nog niet tot stand was gekomen.

Uit het resultaat van dit onderzoek bleek, dat 53% van alle tv-bezitters was aangesloten en dat van de tv-bezitters met belangstelling voor buitenlandse tv-uitzendingen 58% een aansluiting had.

De prognose van de totale vraag kwam dus overeen met de werkelijkheid na drie maanden. Een nadere analyse leidt echter tot de vraag of dit op toeval berust of dat we een dergelijke samenhang altijd zouden mogen verwachten.

Indien namelijk werd nagegaan in hoeverre de aanschafplannen met de werkelijkheid overeenkwamen dan blijkt dat de gerealiseerde vraag ook is gekomen uit de groepen ondervraagden die volgens de septemberenquête niet in de markt waren.

meningen over CAS-abonnement	**alle tv-bezitters**			**tv-bezitters met belangstelling voor buitenl. uitz.**		
	in september opgegeven		**abonn. genomen**	**in september opgegeven**		**abonn. genomen**
	%	**abs.**	**%**	**%**	**abs.**	**%**
zeker abonneren	100	29	66	100	25	64
zeer waarschijnlijk	100	32	56	100	30	60
weet niet, misschien	100	28	46	100	18	44
denk het niet	100	17	65	100	11	63
zeker niet	100	7	14	100	4	25

Aangezien de aantallen ondervraagden in de verschillende categorieën te klein waren om de resultaten verder te analyseren, kon niet worden bepaald welke factoren de verschillen in aanschafplannen en de realisering hiervan hebben beinvloed.

Het lijkt niet onwaarschijnlijk dat de kosten een rol hebben gespeeld. De aanbiedingsprijs in oktober was lager dan de prijs die in de septemberenquête werd genoemd. Dit neemt niet weg, dat het merkwaardig is dat meer dan 60% van de 'denk het niet' categorie tot een CAS abonnement besloot, hetgeen niet lager is dan het percentage afkomstig uit de categorie die zegt zich 'zeker te abonneren'.

Concluderend kan worden gesteld dat de prognose over de totale vraag naar

het CAS-abonnement in belangrijke mate overeenkwam met de werkelijke vraag na drie maanden. Een deel van de abonnementen is hierbij echter gekomen uit de gezinnen waarvan men deze niet verwacht had.

Gezien de belangrijke kosten die dikwijls noodzakelijk zijn voor de ontwikkeling van nieuwe produkten, lijkt het mij van belang voor het marktonderzoek dat wordt gezocht naar prognose-methoden die een meer exact beeld van de markt voor een nieuw produkt kunnen opleveren.

Uit de marktonderzoekwereld

Vaststelling per land van de totale uitgaven voor marktonderzoek

In de afgelopen maanden heeft de Internationale Kamer van Koophandel (I.K.K.) te Parijs zich verdiept in de vraag of het nut kan hebben en zo ja, of het mogelijk is te komen tot vaststelling per land van de nationale uitgaven voor marktonderzoek.

De conclusies waartoe zij is gekomen heeft zij neergelegd in een korte nota, die wij interessant genoeg vonden om haar hier af te drukken.

Zoals men zal zien, beantwoordt de I.K.K. beide onderdelen van de vraag bevestigend. Zij ziet voordelen in een verkrijging of vergroting van inzicht in de omvang van de uitgaven voor marktonderzoek en zij acht het technisch mogelijk het daarvoor nodige cijfermateriaal periodiek bijeen te garen.

Zij verheelt echter geenszins de aan een dergelijke bijeengaring verbonden enorme moeilijkheden en geeft een aantal richtlijnen, welker inachtneming tot een oplossing van deze moeilijkheden zou kunnen bijdragen. Het eindoordeel over de vraag of het, nut en moeite zo goed mogelijk afgemeten, zin zou hebben in Nederland een dergelijk onderzoek aan te vatten, zij aan de lezer overgelaten.

Het pleit voor de I.K.K., dat men na lezing eerder geneigd is te filosoferen over de engte van de pas dan over het uitzicht dat men boven zal hebben.

Organization on uniform lines of surveys of national expenditure on marketing research

Draft recommendations for submission to the 113th session of the Council of the ICC

Foreword

Estimates, so far mostly unofficial, are already appearing in some countries of expenditure on marketing research. Such estimates, if intelligently derived and presented, can be of real benefit to businessmen and should therefore be encouraged but if not soundly based and carefully constructed they can be both misleading to businessmen and damaging to their reputation.

For this reason, the ICC which brings together *inter alia* users and practitioners of marketing research on an international basis and is therefore in a unique position to comment on this matter, has prepared the following document. This makes no attempt to suggest in detail how such market research expenditure surveys should be organized, but merely aims to indicate the basic principles which should be followed in this area.

The value of and basic requirements for such surveys

a. Over the past two decades there has been in many countries a considerable increase in the use of 'business ratios' for inter-firm comparisons. By means of these, businessmen can assess the degree of use and the performance of various resources (e.g. of labour, capital, materials, etc.) within their own firms, against the average for their category of business operation. The ratios are devised to be practical and sensitive indicators of the efficient use of these resources and as such can have considerable advantages for the alert businessman. But this requires that the comparisons are properly constructed and properly understood by the users, otherwise they can be misleading and likely to do more harm than good.

In particular, comparisons in terms of quantities are only valid when similar qualities are being considered; lack of precision and consequent confusion on this aspect can be especially dangerous.

b. With the growth of international trade and the spread of industrial production to the less developed countries, the extension of these comparisons to the international as well as the national plane obviously becomes desirable, and its encouragement is therefore a matter of interest to the International Chamber of Commerce.

c. Primarily, comparisons through 'business ratios' have been used in the field of production, distribution and advertising, but the time may now be appropriate to consider their extension to the field of management resources, including marketing research for which unofficial estimates of some national expenditure have already started to appear.

d. In the area of expenditure on marketing research, such comparisons – although they will not be easy – could help particularly:

– to stimulate interest in commercial and marketing research as a management *tool;* in many countries, both government and private enterprises pay less attention to marketing research than e.g. to research in the physical sciences. This neglect can be serious, since experience shows that the countries the most advanced in their marketing techniques are those where the most thorough marketing research is carried out. A valuable corrective of shortcomings here would be if e.g. educational institutions could be so encouraged by publication of these comparisons to devote more time to the teaching of the techniques and use of marketing research.

– private enterprise to show the public that industry and trade direct considerable efforts towards ensuring that their products or services anticipate and match consumer needs; also that enterprises are making increasing use of marketing research to check the efficiency of their management techniques.

– to give enterprises arguments with which to encourage more governments to take fiscal measures to facilitate marketing research especially in the export field.

Recommendations

For these reasons, the ICC recommends businessmen to consider the development of systems which will provide reliable and effective comparisons of the volume of marketing research. While, to start with, even the formulation within any one country of a methodical system for collecting data will present considerable problems, ultimately – and as research facilities develop towards a more common standard in all countries – the aim should be the provision of uniform and effective international comparisons of expenditure.

At the same time, because of these special difficulties in collecting such data (as has been proved by experience in countries where such studies have already

been attempted) and of ensuring that misleading comparisons are avoided, the ICC also recommends businessmen in making preparations for such surveys to bear in mind that:

a. Marketing research plays an important part in management's role of decision-making: all matters concerning it, including its volume and cost, are carefully guarded secrets which businessmen will not want to risk falling, directly or indirectly, into the hands of their competitors. Therefore arrangements for the confidentiality of data supplied must be exceptionally secure if they are to gain the individual businessman's confidence; also, businessmen before releasing their figures will generally need to be persuaded of the benefits they will gain from co-operation.

b. The actual extraction of the data, the decisions on what to include and what to exclude, may often present difficulties. For example, the boundary between the management costs of interpretation of results (which probably should be included) and those of their implementation in marketing action (which should be excluded) may be difficult to define; similarly 'desk research' important particularly to firms selling industrial goods, often merges with the ordinary marketing and administrative functions of a business so that its true costs are not easily ascertainable.

Therefore, for any such comparative study – whether at national or international level – to be successful, full account of these difficulties must be taken in the framing of its arrangements: instructions on items to be included must be clear and unambigous and based only on what is practical for all participants to provide accurately and without difficulty.

c. Marketing research costs, for work of a comparable kind and quality, vary considerably from country to country, and sometimes between different parts of the same country. This may be partly because of the differing facilities as yet available – these should tend to equate within a few years – and partly because of physical differences (distances, varied populations, etc.) which will always remain. This means that comparisons based solely on expenditure can sometimes be misleading; businessmen arranging studies of this nature should therefore be on the alert for such circumstances and should take account of them in the comparisons shown.

Materials to be included in such surveys

The ICC's role is primarily to indicate the principles which it considers should be observed in such surveys so that the maximum benefit would be gained by businessmen, public bodies and governments in all countries. The details of what can effectively be included, and how proper comparisons may be made, can

only be decided for each country in the light of actual experience in the organizing of such a survey.

Nevertheless some general indications by the ICC of the area to be covered by these surveys in each country seem appropriate since, if the ultimate aim is to achieve international comparability, there must be a major degree of similarity in the content and presentation of the national statistics. If close comparability is not possible world-wide, it might at least be achieved quite soon within groups such as the European Economic Community and EFTA.

The ICC therefore recommends that such surveys should, as an ideal which may not be attainable in all cases, include as many as possible of the following features:

a. Show data separately by industries and/or broad commodity groups – as some accepted international classification.

b. Distinguish between research on home market and for export; also research carried out within the country by a sponsor from another country for purposes of imports. (Normally research should probably be shown under the country where the fieldwork is conducted – but exceptions may be necessary and steps then taken to avoid duplication).

c. Cover all consumer goods and services, and industrial goods.

d. Include any type of research commissioned by enterprises (producers or distributors) in connection with the marketing of any of their goods or services. If possible there should be some broad separation by the main types of research e.g.

demand assessment,
product testing,
attitude and motivation testing,
advertising and communication research,
channels of distribution research,
test marketing studies,
company image

This may not prove easy in practice as many surveys overlap more than one type.

e. Exclude expenditure on basic research done by central or local government bodies (e.g. censuses of population and distribution, town planning research) since the proportions of these relevant to marketing are not at all clear and must vary considerably, but include co-operative research by trade associations, etc. which might if possible be shown under a separate heading. Government sponsored producing and trading bodies should be included under their appropriate industry. Research by academic institutions (e.g. the Marketing Department of Universities), if of direct applicability to particular business prob-

lems, should be included, but not long term 'pure' research or research of a sociological nature with only peripheral interest to particular business problems.

f. Aim to obtain data from 100% of large enterprises, and a sample of smaller enterprises. Contact with both large and small enterprises would probably first need to be made e.g. through research agencies or trade associations (depending on local conditions) but ultimate responsibility for providing the figures, or for permitting them to be provided, should lie always with the body who originally commissioned or requested the research.

In particular, research commissioned by marketing and advertising agencies on behalf of a client, although paid for by the agency, should be included in the client return.

g. Take place at regular intervals.

h. As well as total expenditure (in terms of local currency), seek information on:

1. number of 'interviewer-days' used in the year, divided between skilled and unskilled interviewers.
2. number of 'desk research-days' used in the year.
3. number of persons employed in the marketing and/or market research departments of firms for the purpose of commissioning and applying research provided by agencies.

This might be attempted, at any rate for an experimental period at the inauguration of such surveys, so as to provide an alternative index of volume for international comparisons, independent of wage levels and currency differences.

'Total expenditure' should cover all planning, fieldwork, analysis, interpretation and presentation: 'desk research' should include examination and analysis not only of external information but also similar work with a marketing purpose on internal information (i.e. analysis other than in the normal course of administration or for accountancy purposes): all salaries and overheads for item 3 should be included.

Include, in the presentation of the results, the data translated into the form of some appropriate ratios so as to afford the effective inter-firm (or international) comparisons without which the absolute figures (of industry or national totals) would be of little help to businessmen in assessing their own performances.

The appropriate ratios will probably vary from industry to industry and could best be determined in the light of actual experiment. Value of turnover might be one ratio, while for international comparisons covering all industries or enterprises in each country, the total expenditure might be compared against Gross National Product (as recommended in ICC Brochure 195 for comparisons of advertising expenditure), or e.g. the number of 'interviewer-days' plus 'desk research-days' against total annual working days in each country.

In order to evaluate types of research (discussed under d) in terms of the

amount of money spent on them, separate ratios should be worked out in relation to the particular operation for which a given type of research provides a service, whenever this link is self-evident, e.g. public relations, product development, advertising. This should be attempted in addition to the 'total' ratio which embraces the entire marketing research activity in an enterprise.

Flexibility and ingenuity are required in devising the most revealing ratios. This requirement would, of course, imply that enterprises would have to supply the necessary data for the comparisons (e.g. data on turnover) as well as their research expenditure data.

N.B. In whatever form the figures were published the businessman would still have to use his judgment in applying them to his own needs and should not accept merely superficial comparisons which could be misleading. For example, the manager of a small firm would realise that research expenditure as a % of his turnover might have to be greater than for larger and longer established firms in the same industry; nevertheless, data based largely on returns from these firms would still give him some guidance which would be better than the complete absence of any data at all.

Bodies responsible for sponsoring and conducting such surveys

It is unlikely and indeed undesirable that in any one country the same body would both sponsor and conduct such a survey. Sponsorship would probably best be shared by the users and providers of marketing research – i.e. by the businessmen and the research agencies; both could stand to gain from an effective system, and both should have recognised professional bodies to effect and encourage the co-operation needed.

As to responsibility for conducting the national surveys, one cannot do better than quote the relevant paragraph in ICC Brochure 195 'Promotion of advertising expenditure surveys by class of media on an internationally comparable basis':

'It is one thing, however, to undertake the task of sponsorship (which involves little more than goodwill, willingness to make figures available, and possibly the shouldering of costs which are unlikely to be very great relatively to the importance of the subject) and quite another to undertake the technically very complex work of conducting the survey itself. This sort of work requires the professional skills of expert statisticians and economists: it is not a task for amateurs, nor one that can be treated as a spare-time occupation. Whether it is entrusted to a specially-recruited task-force, to the business administration department of a university, or to some other suitably qualified group is immaterial, provided that the work is put in the hands of people with the appropriate tech-

nical skills and knowledge (of the subject).'

It is even more important in the case of marketing research expenditure, for reasons mentioned above, that the body chosen to conduct the survey should enjoy the full confidence of businessmen both as regards its integrity (in the guarding of confidential figures) and its impartiality.

Whatever the kind of body which may best suit these requirements in any country it is always desirable that the basic form of the survey should be determined by the sponsoring body who will be well acquainted both with what is practical and useful to the businessmen in its country and also with what is required to maximise the international comparability of the results.

Conclusions

The difficulties in launching the first surveys may be considerable, though probably not insurmountable, and as experience is gained so subsequent surveys should become easier, more accurate and more effective – as happened for advertising expenditure.

Broad indications of levels and trends are probably all that are required – at any rate at the first stage – and provided the results contain no flagrant errors of misdirection, minor inaccuracies could be accepted, especially if the limitations in the figures (e.g. difficulties in collection of data and the calculated sampling errors) were clearly recorded.

It is in these terms, and for the reasons set out in this note that the ICC would encourage businessmen to consider the possibility and methods of carrying out such surveys so as to gain not only the benefit of comparisons within their own country but eventually the greater benefits of comparisons against other countries.

NVM-Oriëntatiereis naar Amerika

Eind april 1970 zal een groep van ongeveer twintig N.V.M.-leden naar Amerika vertrekken om zich gedurende een reis van twee à drie weken te oriënteren op het marktonderzoek in Amerika.

Gedurende een eerste bijeenkomst op 19 november jl. hebben de deelnemers overeenstemming bereikt over de volgende te bestuderen onderwerpen:

1. Marktonderzoek in bedrijven en de integratie in de totale marketing en in andere activiteiten. De plaatsbepaling, de functie en de verantwoordelijkheid van de marktonderzoek-afdeling. Welk onderzoek-type, in welke gebieden wordt gebruikt, speciaal de produkt-ontwikkeling en de test-marketing procedures. De voorgestelde bedrijven die voor deze onderwerpen bezocht kunnen worden zijn: Bristol Meyers, New York; Coca-Cola, New York; Procter & Gamble, Cincinnati.
2. Het gebruik van marktonderzoek-data, sales analysis data, etc. voor marketing-modellen, korte en middellange termijn-planning etc.

De voor deze onderwerpen voorgestelde bedrijven zijn: Bristol Meyers, New York; Coca-Cola, New York; Procter & Gamble, Cincinnati; Marketing Science Institute, Boston; Massachusetts Institute of Technology (M.I.T.), Boston; Survey Research Center van de University of Michigan, Ann Arbor.

3. Communications research, advertising research en media selectie. Voor deze onderwerpen-groep worden voorgesteld de volgende bedrijven: Advertising Research Foundation, New York; Young & Rubicam, New York; J. Walter Thompson, New York; Gallup & Robinson; Princeton, N.J.; McGraw-Hill, New York.
4. Social research en macro-economic research. Voor deze onderwerpen worden voorgesteld: Survey Research Center (Katona) van de University of Michigan, Ann Arbor.
5. Distributie-ontwikkelingen, shopping centers en assortiments-bepaling.

Hiervoor wordt als bedrijf gesuggereerd het Supermarket Institute, Chicago.

6. Opleiding voor marktonderzoeken.
7. Marktonderzoekbureaus.

Voor deze reis die ongeveer twee à twee en een half duizend gulden gaat kosten, inclusief vervoer naar Amerika en binnenlands vervoer naar de verschillende bedrijven, overnachting, kunnen zich nog enkele deelnemers opgeven. Men wordt verzocht zich te wenden tot het secretariaat voor nadere informaties en voor eventuele opgaven.

Literatuurbespreking

MARKET RESEARCH ABSTRACTS

Het aantal tijdschriften waarin marktonderzoekonderwerpen aan de orde komen is reeds zo groot dat er geruime tijd geleden de noodzaak gevoeld werd om de toegankelijkheid hiertoe te vergemakkelijken.

In dit verband wijzen wij de lezers van dit tijdschrift op het bestaan van Market Research Abstracts, een uitgave van The Market Research Society.

Artikelen uit de volgende tijdschriften worden hierin gerubriceerd, terwijl een korte inhoud van het betreffende artikel wordt weergegeven.

Research

Journal of the Market Research Society
MRS Conference
IMRA Journal
Journal of Advertising Research
Journal of Marketing Research
Metra
Operational Research Quarterly
Operations Research

Statistical

American Statistician
Bulletin of the Oxford University Institute of Economics & Statistics
Journal of the American Statistical Association
Journal of the Royal Statistical Society Series A (General)
Series C (Applied Statistics)
The Statistician

Psychological/Sociological

American Journal of Sociology
Behavioral Science
British Journal of Psychology
British Journal of Mathematical & Statistical Psychology
British Journal of Social & Clinical Psychology
British Journal of Sociology
Educational & Psychological Measurement
Human Relations
Journal of Abnormal Psychology
Journal of Personality & Social Psychology
Journal of Applied Psychology
Journal of Social Issues
Occupational Psychology
Psychological Bulletin
Psychological Review

Economics/Marketing/Advertising/Business Management

ADMAP
Advertising Quarterly
British Journal of Marketing
National Institute Economic Review
Harvard Business Review
Journal of Industrial Economics
Journal of Marketing
Management Decision
Marketing Communications
Public Opinion Quarterly
Revue Française du Marketing

De volgende rubrieken worden gehanteerd:

Section I	:	Survey Techniques
Section II	:	Statistics, Models and Forecasting
Section III	:	Attitude and Behaviour Research
Section IV	:	Psychology
Section V	:	Advertising and Media Research

Section VI : Applications of Research
Section VII : Industrial Marketing Research
Section VIII : Industrial Marketing Research

Daarnaast is er een trefwoordenregister, zodat men bijvoorbeeld bij interesse in het onderwerp 'product testing' dit trefwoord opzoekt, waarna men verwezen wordt naar artikelen over dit onderwerp.

Market Research Abstracts is een halfjaarlijkse uitgave, waarmee gestart is in 1963.

De laatste uitgave bestrijkt de periode van januari-juni 1968.

De kosten bedragen £ 12.15 s voor 3 jaar.

JOURNAL OF MARKETING RESEARCH

August 1969

Management Control of Marketing Research
James H. Myers and *A. Coskun Samli*

Determinants of Buyer Response to Physical Distribution Service
Ronald P. Willett and *P. Ronald Stephenson*

Interest Patterns and Product Preferences: An Exploratory Analysis
Allan L. Pennington and *Robert A. Peterson*

A Simultaneous Equation Regression Study of Advertising and Sales of Cigarettes
Frank M. Bass

The Effect of Advertising on Liquor Brand Sales
Julian L. Simon

Advertising Effectiveness and Risk in the Consumer Decision Process
Jeffrey A. Barach

Consumer's Purchase Decision Process: Stochastic Models
Tanniru R. Rao

Multidimensional Scaling: An Introduction and Comparison of Nonmetric Unfolding Techniques
Paul E. Green and *Frank J. Carmone*

Conditional Trend Analysis: A Model that Allows for Nonusers
Donald G. Morrison

Experimental Analysis of Theory of Promotion at Point of Consumption
Robin T. Peterson

Canonical Analysis of Relations between Socioeconomic Risk and Personal Influence in Purchase Decisions
Michael Perry and *B. Curtis Hamm*

Computer Applications

A Series Approximation for Negative Binominal Parameter Estimation
Donald G. Morrison

SYMAP: Computer Graphics for Marketing Management
Jack S. Wolf

Communications

A Note on Proximity Measures and Cluster Analysis
Paul E. Green and *Vithala R. Rao*

Similarity of Ghetto and Nonghetto Food Costs
Burton H. Marcus

Semantic Generalization in Formation of Consumer Attitudes
Hope S. Roman

Underground Testing
Neil Bruce Holbert

EDITORIAL: An Editor's Farewell
Robert Ferber

BOOK REVIEWS

JOURNAL OF THE MARKET RESEARCH SOCIETY

Volume II number 4

Towards an integrated theory of consumer behaviour
A. S. C. Ehrenberg

Towards a validation of the poster audience model for Greater London
B. D. Copland

Self-concept and brand preference: an empirical application of multidimensional scaling
Paul E. Green, Arun Maheshwari and *Vithala R. Rao*

Order effects in taste tests
D. Berdy

Conference report

ESOMAR 1969: Amsterdam

Publications

Book reviews
Colin R. Bell: Middle class families
P. Sampson
J. A. Banks (editor): Studies in British society
L. F. Collins
Simon Yudkin and *Anthea Holme:* Working mothers and their children
Hannah Gavron: The captive wife
Pat Williams: Working wonders
Anne Ward
Ann Roush: Selling to North American stores
A. Wolfe
Robert J. Alsegg: Researching the European markets
A. Turner
Gordon Wills (editor): Sources of UK marketing information
Susan Johnson
I. G. Anderson (editor): Marketing and management: a world register of organisations
Susan Johnson
Richard Graves and *Allan Campbell:* Creating customers
E. Adler
Martin Christopher: A cluster analysis of towns in England and Wales according to their suitability for test market locations
C. Holmes
Elizabeth Beazley: Designed for recreation
J. G. Field
Ruari McLean: Magazine design
J. G. Field

HMSO publications received

Publications received

Medewerkers

P. Camfferman (47) is kwaliteitstechnicus bij Philips' Gloeilampenfabrieken. Hij volgde een electrotechnische opleiding. Hij schreef enkele artikelen in medische tijdschriften.

C. P. L. Commissaris (40) is als Research Engineer verbonden aan de Nijverheidsorganisatie TNO. Publiceert geregeld in tijdschriften op technisch gebied.

C. J. Dollé (41) is verbonden aan Philips' Gloeilampenfabrieken. Drs. Dollé studeerde aan de Universiteit van Amsterdam (psychologie). Hij is belast met 'product research' waarbij het zwaartepunt op de methodologie ligt. Heeft enkele publikaties op zijn naam staan waaronder een monografie in de serie Documentatie onderzoektechnieken van de Nederlandse Vereniging van Marktonderzoekers.

A. W. J. de Gee (36) is afdelingsleider bij de Nijverheidsorganisatie TNO. Ir. de Gee studeerde in Delft (technoloog richting T). Hij heeft een groot aantal artikelen geschreven voornamelijk op het gebied van technisch onderzoek.

O. Grosch (42) is hoofd van de Technische Afdeling van de Nederlandse Consumentenbond in Den Haag. Ir. Grosch, die in Delft (elektrotechniek) en in Rotterdam studeerde, was onder meer werkzaam bij de NIT en bij Unilever Vlaardingen.

A. W. Jongert (47) is hoofd Extern Marktonderzoek van de PTT. Drs. Jongert studeerde in Utrecht (economische geografie). Was o.m. werkzaam bij het NIPO.

E.N. Koster (45) is Service and Personnel Manager bij het Instituut voor Huishoudelijk Onderzoek (Unilever). Drs. Koster studeerde in Rotterdam (economie). Voor hij bij het IHO in dienst trad, was hij o.m. werkzaam op Curaçao bij de Meduro Bank. Drs. Koster publiceerde reeds eerder samen met de heer Melt in het Marktonderzoek Kwartaalschrift.

W. van Vliet (42) is Market Research Manager voor Europa van de Mars Chocolade Fabriek. Daarvoor was hij o.a. werkzaam als verzekeringswiskundige en werkte hij bij Attwood Statistics.