

Rotte appels?

Controle op kwaliteit van antwoordgedrag in het Intomart GfK online panel

*Lex van Meurs, Robert van Ossenbruggen
en Liesbeth Nekkers*

In: Bronner, A.E et all (eds) *Ontwikkelingen in het marktonderzoek: Jaarboek MarktOnderzoekAssociatie, dl. 34, 2009*. Haarlem: Spaar^{en}Hout

Samenvatting

Ter verhoging van de kwaliteit van online onderzoek heeft Intomart GfK in 2007 een kwaliteitssysteem geïntroduceerd ('FLAGS') dat frauderende respondenten automatisch opspoor op basis van invulduur, responsgedrag op stellingenblokken ("straight lining"), en het beantwoorden van open vragen en screeningsvragen.

In dit artikel worden de resultaten van 86 onderzoeken uit het Intomart GfK online panel besproken. Op basis van onze definities komt in dit panel verdacht antwoordgedrag maar in 4% van alle gevallen voor en hangt niet of nauwelijks samen met kenmerken van de respondenten als leeftijd of opleiding.

Met FLAGS kunnen ook vragenlijsten worden gedetecteerd die dubieus respondentengedrag uitlokken. Dit bleken vooral vragenlijsten te zijn die als minder interessant, te lang of te saai werden beoordeeld. De meest effectieve manier om frauduleus invullen van online onderzoek te voorkomen blijkt de kwaliteit van de vragenlijsten te zijn.

Met het kwaliteitssysteem lopen frauderende respondenten automatisch tegen de lamp en worden concrete handvatten geboden om samen met opdrachtgevers dubieus respondentengedrag te voorkomen door verdere verbetering van de kwaliteit van vragenlijsten.

Trefwoorden: *online panels, panelfraude, data integriteit, datakwaliteit, vragenlijstconstructie*

1. Inleiding: frauduleus invulgedrag kan worden beperkt met betere vragenlijsten

Om de kwaliteit van onderzoek door middel van internetvragenlijsten verder te verhogen, heeft Intomart GfK in november 2007 het FLAGS-systeem geïntroduceerd. Met behulp van dit systeem wordt op systematische wijze de kwaliteit van online verzamelde onderzoeksdata gecontroleerd. Hierdoor kunnen niet alleen frauderende respondenten makkelijker worden

opgespoord, maar worden ook vragenlijsten gedetecteerd die dubieus respondentengedrag uitlokken.

FLAGS is tot stand gekomen na uitgebreid kwaliteitsonderzoek binnen het Intomart GfK online panel. Hiervoor werden 86 projecten geëvalueerd. In dit artikel worden de resultaten van dit onderzoek en de opzet van de methode samengevat. Deze aanpak is eerder gepresenteerd op het ESOMAR panel congres in oktober 2007 (van Meurs, Ossenbruggen and Nekkers, 2007).

Met behulp van het kwaliteitssysteem wordt tijdens het veldwerk op systematische wijze gecontroleerd op bewust slordig invullen van vragenlijsten op basis van invulduur, responsgedrag op stellingenblokken (“straight lining”) en het beantwoorden van open vragen en zogenaamde screeningsvragen.

Door gebruik te maken van FLAGS kunnen te slordig ingevulde vragenlijsten en eventuele frauderende respondenten eenvoudig en snel worden opgespoord. Bij de in artikel beschreven test bleken minder dan 4% van alle responses als verdacht aangemerkt. Op basis van deze resultaten uit één panel vermoeden we dat dubieus respondentengedrag in een optimaal gemanaged panel maar incidenteel voorkomt. Verdacht antwoordgedrag hangt niet samen met kenmerken van de respondenten als leeftijd of opleiding.

Met het kwaliteitssysteem kunnen ook vragenlijsten worden gedetecteerd die dubieus respondentengedrag uitlokken. Onderzoeken waarbij relatief veel verdachte antwoorden werden aangetroffen bleken vooral vragenlijsten te zijn die door respondenten als minder interessant, te lang of te saai werden beoordeeld. Door datakwaliteit systematisch en automatisch te controleren krijgen onderzoekers en opdrachtgevers concrete handvatten om de kwaliteit van online onderzoek te verhogen. De meest effectieve manier om frauduleus invullen van online onderzoek te voorkomen blijkt de kwaliteit van de vragenlijsten te zijn.

2. Responsekwaliteit bij online onderzoek

2.1 De ontwikkeling van een kwaliteitssysteem

Online onderzoek heeft de afgelopen jaren een wel heel stormachtige groei doorgemaakt. De aandacht voor de kwaliteit van de data die met deze methodiek worden verzameld is helaas niet altijd in gelijke mate meegegroeid.

In dit artikel beperken we ons tot online onderzoek dat wordt gehouden in panels. Bij het samenstellen van steekproeven uit access panels bestaat het risico op beroepsrespondenten en “zware” internetgebruikers. Om dit te voorkomen is de manier waarop een online panel wordt opgebouwd en beheerd van groot belang. In dit artikel worden de resultaten van een nieuw kwaliteitscontrolesysteem beschreven aan de hand van het Intomart GfK Online Panel. Dit

panel is grotendeels verkregen uit telefonisch onderzoek gebaseerd op verse landelijk representatieve steekproeven waarmee een juiste afspiegeling sterk wordt bevorderd. Hoewel de belangrijkste conclusies voor de gehele branche relevant zijn, kunnen de resultaten van dit onderzoek niet zonder meer worden gegeneraliseerd naar andere Nederlandse online panels.

Online dataverzameling vereist specifieke kwaliteitscontroles. Anders dan bij face-to-face of telefonisch onderzoek is er geen sprake van directe participatie in de dataverzameling door een waarnemer (interviewer). De respondent wordt ten tijde van het interview niet op de vingers gekeken. Dit vrijwaart de onderzoeksgegevens van een vertekening door sociaal gewenste antwoorden, maar een keerzijde is dat controles achteraf scherper dienen te zijn. Door Intomart GfK is dan ook een kwaliteitssysteem ontwikkeld dat bij alle online onderzoeken op ons panel meeloopt: “FLAGS: the GfK Survey and Panellist Quality Filter”.

Dit systeem heeft twee doelen:

1. mogelijk niet valide interviews in het onderzoek te signaleren;
2. panelleden van mogelijk twijfelachtige kwaliteit te signaleren en uit het online panel te verwijderen.

Hiertoe maakt het systeem gebruik van een aantal parameters dat tijdens elk interview wordt berekend. Het betreft hier onder meer:

- De gemiddelde tijd die een respondent er over doet om een vraag te beantwoorden. Hierbij gaan we er vanuit dat er een minimale benodigde tijd is om een vraag te kunnen lezen en serieus te beantwoorden;
- Open antwoorden: het percentage open vragen dat de respondent niet beantwoordt. Deze parameter is een maat voor de “luiheid” van de respondent;
- Straight-lining: in hoeverre een respondent op blokken van vragen steeds precies hetzelfde antwoord geeft;

De resultaten op deze parameters worden voor elk onderzoek bijgehouden, zodat panelleden met structureel twijfelachtige resultaten eenvoudig kunnen worden opgespoord en uit het panel worden verwijderd. De kwaliteit van het panel neemt hierdoor continue toe.

2.2 Satisficing: optimaal of slordig invulgedrag

Om een correct antwoord te geven op een vraag dient een respondent vier cognitieve stappen te doorlopen (zie bijvoorbeeld Sudman, Bradburn & Schwarz, 1996 en Tourangeau, Rips & Rasinski, 2000). Een respondent wordt verondersteld:

1. de vraag te begrijpen;
2. zich de informatie te herinneren die nodig is voor een antwoord;
3. deze informatie af te wegen;
4. een antwoordcategorie te selecteren.

We noemen een response optimaal (“*optimizing*”) als de respondent deze stappen nauwgezet en zonder ongewenste beïnvloeding doorloopt (zie bijvoorbeeld Simon, 1957). Maar elk van deze vier stappen kan behoorlijk ingewikkeld zijn, waarbij een grote mate van cognitieve inspanning is vereist. Een respondent kan het zich makkelijker maken door de vier cognitieve stappen wat sneller te doorlopen. Dit kan een antwoord geven dat nog wel acceptabel genoemd kan worden, maar niet optimaal is. Dergelijk slordig invulgedrag wordt ook wel *satisficing* genoemd. Krosnick (1999) maakt onderscheid tussen een zwakke en sterke mate van satisficing. Zwakke satisficing vindt plaats wanneer een respondent niet nauwgezet genoeg de vier stappen doorloopt, bijvoorbeeld niet voldoende tijd en inspanning besteedt aan het begrijpen van de vraag, het zich herinneren van de benodigde informatie, het afwegen hiervan of het selecteren van een antwoordcategorie. In veel gevallen hoeven de gevolgen hiervan voor de validiteit van de antwoorden nog niet substantieel te zijn.

Een sterkere mate van *satisficing* vindt plaats wanneer een respondent de vragen in een extremere mate slordig invult, bijvoorbeeld door de vraag zo snel mogelijk door te nemen en zonder veel nadenken de eerste voor de hand liggende antwoordcategorie kiest. In vragenlijstonderzoek kan dit leiden tot een vertekening door middel van relatief veel positieve, neutrale of stereotypische antwoorden. Dit effect kan worden versterkt door moeilijke vragen of door een relatief minder gemotiveerde respondent. Voor respondenten hiervan de schuld te geven zouden onderzoekers eerst de verantwoordelijkheid moeten nemen om dergelijk ongewenst invulgedrag te voorkomen door de taak van respondenten te vereenvoudigen door middel van bijvoorbeeld een beter ontwerp van vragenlijsten.

2.3 Frauduleus antwoordgedrag

Het slordig en gemakzuchtig invullen van vragenlijsten vormt een bedreiging voor de validiteit van de onderzoeksuitkomsten. Het wordt erger wanneer respondenten bij het invullen van de vragen één of meer van de vier noodzakelijke cognitieve stappen overslaan, bijvoorbeeld een vraag invullen zonder deze te lezen of na te denken (zie bijvoorbeeld Smith & Hofma Brown, 2005; Knapton & Garlick, 2006). Sinds kort lijken aanbieders van onderzoekspanels de bedreiging van dergelijk frauduleus antwoordgedrag serieus te nemen. Veel panel providers stellen hun gebruikers gerust dat hun panel is gevrijwaard van frauderende respondenten of dat hier tenminste goed op wordt gelet. Een goed voorbeeld hiervan is GMI’s ‘PureSample’ initiatief waarbij panel providers en hun gebruikers de emailadressen van panelleden kunnen laten controleren op als frauderend bekend staande personen. Fraudeurs kunnen ook op een zwarte lijst worden gezet in de PureSample database. Met inmiddels ruim 1 miljoen emailadressen op deze snel groeiende zwarte lijst claimt GMI een belangrijke kwaliteitsimpuls te kunnen geven aan het online onderzoek.

Dit is een duidelijk voorbeeld van het leggen van de schuld bij de respondent, zonder dat we als onderzoekers precies begrijpen wat er aan de hand is en hoe ongewenst gedrag kan

worden voorkomen. Smith en Hofma Brown (2006, p. 20) zeggen hierover: “With so many possible sources of error (...) researchers, suppliers, clients and the industry need to be careful to react to crucial challenges to panel and research quality without over-reacting.” Het is nog maar de vraag of je kwaadwillende fraudeurs tegenhoudt met dergelijke zwarte lijsten. Als we als onderzoekers onze panelleden als potentiële fraudeurs gaan behandelen, vormen we zelf wellicht een grotere bedreiging voor de toekomst van de onderzoeksbranche dan de paar fraudeurs die hiermee uit onderzoek geweerd kan worden.

We geloven dat we beter stappen kunnen ondernemen om respondenten niet als vragenlijstrobots te behandelen, maar bijvoorbeeld te streven naar het bekorten van vragenlijsten, meer afwisseling in vraagvormen en onderwerpen te bieden en respondenten te vragen naar feedback over vragenlijsten. Comley (2006) geeft suggesties voor een aantal simpele maar effectieve maatregelen om de relatie met respondenten te verbeteren en ze een betere vragenlijstervaring te bieden. Daarnaast dient de onderzoeksbranche tot overeenstemming te komen over wat wel en niet acceptabel antwoordgedrag is en met welke indicatoren dit kan worden gesignaleerd. Met dit artikel willen wij aan dat proces een bijdrage leveren.

2.4 Professionele respondenten

Bij online panel onderzoek is er een grotere kans op extreem slordig of frauduleus antwoordgedrag dan bij andere veldwerkmethoden. Een belangrijk aspect van in ieder geval een belangrijk deel van de online panels is dat ze sterk gericht zijn op beloningen. De onderzoeksbranche maakt zich zorgen over ‘professionele respondenten’ die vooral zijn gemotiveerd om zo snel en gemakkelijk mogelijk geld te verdienen door middel van onderzoek. De onderstaande voorbeelden van websites waarmee respondenten worden geworven geven hiervan een verontrustend beeld:

- “Discover How To Consistently Receive Huge Checks At Your Doorstep From Big Corporations Who Are Dying To Know Your Opinion!” (www.InstantPaidSurveys.com)
- “Earn \$10 in 30 minutes. We want to help you get started in surveys. Receive \$10 cash by tomorrow, when you register at 20 of the Top 25 Surveys. \$\$ Have your money in 24 hours \$\$” (www.thesurveypro.com)
- “Make \$10-\$150 An Hour Taking Fun, Exciting Paid Surveys” (www.yellowsurveys.com)
- “Paid surveys are perfect for all students, stay-at-home moms, and anyone else that needs to make extra money ... and it's FREE!” (www.surveyclub.com)

Het is niet direct duidelijk welke bedreiging ‘professionele respondenten’ vormen voor panelonderzoek. Harris Interactive (Smith & Hofma Brown, 2005) identificeerde ‘onoplettende’ en ‘frauderende’ respondenten, maar concludeerden dat hoewel dergelijke respondenten kunnen worden geïdentificeerd, ze geen werkelijke bedreiging vormen voor de

kwaliteit van de onderzoeksresultaten. Knapton en Garlick (2006) tonen echter aan dat sommige groepen respondenten daadwerkelijk de datavaliditeit kunnen bedreigen. Willems, van Ossenbruggen en Vonk (2006) tonen aan dat de verhouding van ‘gemakzuchtige’, ‘professionele’ en ‘loyale’ panelliden varieert tussen verschillende onderzoekspanels waarmee verschillen in onderzoeksresultaten deels verklaard kunnen worden. Waarschijnlijk hangen deze verschillen in panelsamenstelling samen met panel management factoren zoals de wervingsmethode en het beloningssysteem.

Naast de al genoemde afwezigheid van interviewers is een ander kenmerk van online onderzoek dat het opdrachtgevers de mogelijkheid geeft om voor minder budget meer onderzoek uit te voeren. Dit heeft een sterke toename van het aantal onderzoeken tot gevolg gehad. Omdat de groei van panels is achterbleven op deze toename, is de onderzoeksdruk per respondent toegenomen. We achten het waarschijnlijk dat met de toegenomen onderzoeksdruk sommige respondenten hun toevlucht zullen zoeken in het sneller en makkelijker invullen van vragenlijsten. Het risico op slordig of frauduleus antwoordgedrag bij online onderzoek neemt hierdoor toe.

2.5 Drie niveaus van responsekwaliteit

Hoewel een aantal onderzoeken ons bewust hebben gemaakt van de verschillen in de kwaliteit van antwoordgedrag en hiervoor indicatoren zijn geïdentificeerd, is er nog weinig of geen onderzoek beschikbaar naar de gevolgen hiervan. In dit onderzoek met als case study het Intomart GfK online panel is onderscheid gemaakt naar drie verschillende niveaus van antwoordgedrag.

Ten eerste beschouwen we incidenteel slecht antwoordgedrag. Dit kan worden veroorzaakt doordat een panellid gewoon haar dag niet had en de volgende keer weer wel beter haar best doet. In dit geval kan de kwaliteit en validiteit van het betreffende onderzoek zijn geschaad, maar er is geen reden deze respondent uit verder onderzoek te weren.

Ten tweede kijken we of bepaald onderzoek slecht antwoordgedrag oproept. Een vragenlijst kan saai of onduidelijk geformuleerd zijn, met teveel herhaling of andere kenmerken die het gewenst antwoordgedrag belemmeren. In dit geval dienen we de fout bij onszelf als onderzoekers te zoeken.

Ten derde zouden we slecht antwoordgedrag systematisch willen volgen, waarbij we bij respondenten die een vragenlijst slordig of frauduleus hebben ingevuld controleren hoe deze andere vragenlijsten invullen. Bij een systematisch onvoldoende antwoordgedrag zou een dergelijke respondent uit het panel kunnen worden geweerd indien we kunnen aantonen dat deze respondenten de validiteit en kwaliteit van de onderzoeksuitkomsten bedreigen. Helaas waren bij het publiceren van dit artikel voor dergelijke analyses nog onvoldoende observaties beschikbaar.

2.6 Doel van het onderzoek

Om de meest effectieve manier vast te stellen om slecht antwoordgedrag testen we in dit onderzoek een aantal verschillende indicatoren die de afgelopen jaren door diverse auteurs zijn gesuggereerd. We gaan hierbij op zoek naar respondenten met ondermaats antwoordgedrag. Wie zijn deze zwarte schapen en op welke wijze beïnvloeden zij de onderzoeksuitkomsten? We zullen ook kijken naar vragenlijsten die dergelijk gedrag uitlokken: bij welk soort onderzoek komt slecht antwoordgedrag relatief het meest voor? Op welke wijze kunnen deze onderzoeken worden verbeterd om herhaling te voorkomen?

3. Het Intomart GfK online panel

Sinds 1999 voert Intomart GfK onderzoek uit via een online consumentenpanel. Op dit moment bestaat dit panel uit 130.000 actieve leden van 13 jaar en ouder. Panelleden worden gemiddeld tweemaal per maand via een e-mail uitgenodigd voor het invullen van een online vragenlijst. Zij klikken daarbij op een link in de e-mail waardoor ze automatisch in de vragenlijst terecht komen.

De onderzoeksgegevens in dit artikel betreffen het Intomart GfK online panel. Uit het NOPVO onderzoek bleek reeds dat dit panel zich op positieve wijze onderscheidt van andere Nederlandse online panels (Willems, van Ossenbruggen en Vonk, 2006). Om de uitkomsten in dit artikel te kunnen plaatsen wordt in deze paragraaf als achtergrondinformatie beschreven hoe het Intomart GfK online panel wordt geworven, op welke wijze het panel management wordt uitgevoerd, het incentive systeem, de wijze van steekproeftrekking en de response die het Intomart GfK online panel wordt behaald.

Werving

Een belangrijke vereiste voor het kunnen trekken van representatieve online steekproeven is het beschikken over een zo representatief mogelijke wervingsmethodiek. Voor het Intomart GfK online panel wordt weinig gebruik van spontane aanmeldingen, maar is het overgrote deel van ons panel geworven vanuit zelf uitgevoerd landelijk representatief telefonisch, schriftelijk en face to face onderzoek. Via andere methoden (zoals bijvoorbeeld gerichte werving via nieuwsbrieven van web portals) wordt geprobeerd groepen die onvoldoende worden bereikt alsnog voor deelname te interesseren. In tegenstelling tot panels die online door middel van bijvoorbeeld pop-up schermen worden geworven, vertegenwoordigt het Intomart GfK Online Panel in hoge mate alle soorten Internetgebruikers, dus ook diegenen die er relatief weinig gebruik van maken. Iedere internetgebruiker heeft een even grote kans om in het panel terecht te komen. Heavy internet users zijn in het Intomart GfK online panel dan ook niet zo sterk oververtegenwoordigd als in internet access panels waarbij werving via internet plaatsvindt. Bovendien is de meerderheid van het panel slechts bij het Intomart GfK panel aangemeld en neemt niet deel in panels van andere onderzoeksbureaus.

Panelmanagement en beloning

Intomart GfK heeft sinds de jaren 60 ervaring met het beheren van panels, zoals bijvoorbeeld het Kijkerspanel (sinds 1965) en het Radiopanel (sinds 1967). Intomart GfK beschikt over een speciale afdeling Panelbeheer, waaronder ook een Helpdesk valt die via e-mail en telefoon voor respondenten te bereiken valt. Deze registreert en update actief de achtergrondgegevens van de panelleden. Als bij de jaarlijkse, voor respondenten verplichte, update van deze gegevens geen reactie volgt, wordt de betreffende deelnemer uitgesloten van toekomstige deelname. Er vindt ook voortdurend instroom van nieuwe panelleden plaats, met jaarlijks een verversingsgraad van 20-25%.

De respondent krijgt voor zijn/haar deelname een vergoeding in de vorm van het toevoegen van een bedrag aan zijn/haar saldo. Als dit saldo een bepaalde hoogte heeft bereikt kan men ervoor kiezen het te besteden aan een goed doel, er een cadeaubon voor te krijgen of het bedrag online te besteden bij Bol.com. De hoogte van de incentive wordt vastgesteld op basis van de gemiddelde duur van de vragenlijst. Voor een vragenlijst van 5 minuten wordt een incentive van € 1,00 gegeven, voor een vragenlijst van 15 minuten € 2,25. Alleen bij screeningsonderzoeken wordt een loterij gehanteerd, waarbij enkele respondenten een prijs ontvangen. Niettemin wordt bij deze onderzoeken een respons van gemiddeld 83% behaald.

Steekproeven trekken

Een groot aantal achtergrondkenmerken van de panelleden is vooraf bekend. Dat betekent dat op basis van deze gegevens een gestratificeerde steekproef naar achtergrondkenmerken getrokken kan worden om ervoor te zorgen dat alle 'soorten' burgers in voldoende mate in de netto respons vertegenwoordigd zullen zijn. Kenmerken waarover wij vooraf al kunnen beschikken zijn onder andere ook de variabelen die van belang zijn voor de steekproeftrekking en weging in het kader van de campagne-effectmetingen: o.a. geslacht, leeftijd, opleiding, huishoudgrootte, regio en internetgebruik. Het op deze manier trekken van steekproeven in combinatie met een hoge response (zie hieronder) zorgt ervoor dat voor weging achteraf slechts in beperkte mate noodzaak bestaat.

Response

De gemiddelde response in het Intomart GfK online panel is 76% (percentage van de uitgenodigde respondenten die de vragenlijst volledig invullen). Bij onderzoeken met veel maatschappelijke relevantie (bijvoorbeeld leefbaarheid & veiligheid) of onderzoeken met veel beeld- en geluidsmateriaal pakt de respons over het algemeen hoger uit.

Om een dergelijk relatief hoge respons te kunnen blijven garanderen wordt de kwaliteit van de vragenlijsten bewaakt. Irritatie over de opzet van vragenlijsten kan namelijk leiden tot uitval of verminderde responsbereidheid. Daarnaast bewaakt Intomart GfK de belasting van respondenten: het mag niet te veel, maar ook niet te weinig zijn. Ook moet de respondent

enige tijd krijgen om een lijst in te vullen en ook aan de vragenlijst herinnerd worden, zodat in de respons niet alleen deelnemers opgenomen zijn die bij wijze van spreken 24 uur per dag online zijn.

Responsebelasting

Als een panellid te weinig wordt uitgenodigd kan de betrokkenheid dalen, maar bij overbelasting kan de kwaliteit onder druk komen te staan. Gemiddeld wordt een respondent twee maal per maand uitgenodigd voor deelname aan een onderzoek.

Van de online panelleden wordt 10% in een halfjaar tijd 16 keer of vaker wordt uitgenodigd voor deelname. Ongeveer de helft wordt 11-15 keer benaderd. De resterende 40% wordt in een half jaar tijd 6-10 keer uitgenodigd. Het gemiddelde levert de gewenste twee uitnodigingen per maand op.

Evaluatievragen

Aan het eind van iedere online onderzoek wordt 20% van de steekproef gevraagd de vragenlijst te beoordelen op basis van een viertal vaste evaluatievragen: de waardering door middel van een rapportcijfer, de mate van interesse in het onderwerp en de begrijpelijkheid, de duur en de mate van herhaling in de vragenlijst. In paragraaf 6.3 worden deze gegevens nader beschreven.

4. Veldwerk

Dit onderzoek is gebaseerd op 86 random gekozen onderzoeken die van januari t/m augustus 2007 in het Intomart GfK online panel werden gehouden. De gemiddelde netto steekproefgrootte van de 86 onderzoeken was $n=1.360$ met totaal 59.089 observaties. De kleinste steekproef betrof 81 respondenten, de grootste was een screeningsvragenlijst van $n=69.917$. In totaal participeerden 38.969 respondenten aan één of meer onderzoeken. De gemiddelde vragenlijst duur was 8 minuten. De langste vragenlijst duurde 29 minuten. Er waren ook drie onderzoeken met een gemiddelde invulduur van een minuut of minder.

De 86 onderzoeken betroffen in meer dan de helft van de gevallen marketing of consumenten onderzoek (48 cases), 18 betroffen onderzoek voor de overheid of een non profit opdrachtgever, en de overige 20 onderzoeken werden uitgevoerd ten behoeve van media exploitanten.

5. Indicatoren voor de kwaliteit van antwoordgedrag

Voor het identificeren van de kwaliteit van antwoorden in vragenlijstonderzoek kunnen we onderscheid maken naar de volgende soorten antwoordgedrag:

- het correct identificeren van de respondent (wordt de vragenlijst niet door iemand anders beantwoord);
- genoeg tijd nemen voor het lezen van de vraag en de antwoordcategorieën;
- genoeg tijd nemen voor het beantwoorden van de vraag (zonder te overrationaliseren);
- het in voldoende mate beantwoorden van open vragen (de antwoorden moeten op zijn minst geclassificeerd kunnen worden);
- het geven van consistente antwoorden.

We hebben geëxperimenteerd met de volgende criteria:

- invulduur
- straight lining
- open antwoorden
- weet niet
- screeningsvragen
- consistentiechecks

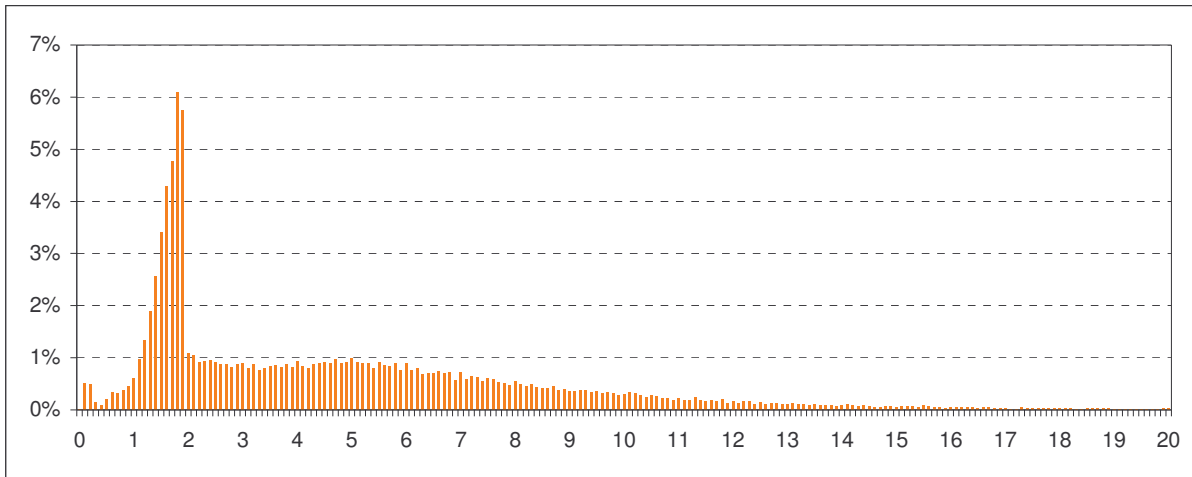
Deze criteria worden hieronder nader beschreven.

5.1 Invulduur

Het lezen van een vraag en de antwoordcategorieën kost een bepaalde tijd. Het sneller beantwoorden van de vraag is verdacht. Bij het berekenen van de minimale invulduur van een vragenlijst dient rekening gehouden te worden met de routing en/of het aantal schermen die aan elke respondent worden getoond. Respondenten met een extreem lange invulduur zijn verwijderd uit de analyse, er rekening mee houdend dat respondenten tijdens een vragenlijst soms even pauzeren.

In figuur 1 wordt de frequentieverdeling gegeven van de gemiddelde invulduur per vraag (het aantal gekozen antwoordcategorieën). Er is een scherpe daling van de gemiddelde invulduur na 2 seconden. We vermoeden dat de extreem langere invulduur is veroorzaakt door respondenten die tijdens het invullen van de vragenlijsten even pauzeerden zonder de online verbinding met de vragenlijstserver te verbreken.

Ten behoeve van onze analyses is 0,4 seconde per antwoord bij minimaal 30 gegeven antwoorden gekozen als (arbitrair) criterium voor de minimaal acceptabele invulduur (exclusief vragen die door routing niet zijn getoond).



Figuur 1: Frequentieverdeling van de gemiddelde invulduur per vraag, $n = 52.726$ (Selectie: minimaal 30 responses per respondent, maximum invulduur 45 minuten)

5.2 Straight lining

Straight lining (het kiezen van steeds hetzelfde antwoord) kan een manier zijn om het invullen van een vragenlijst te versnellen, maar het zou ook de werkelijke mening van een respondent kunnen zijn. Het kan ook wijzen op een slecht samengestelde vragenlijst. Ten behoeve van ons onderzoek is straight lining als volgt geoperationaliseerd:

Er is sprake van straight lining in een vragenlijst wanneer er geen variantie in de antwoorden is bij twee of meer groepen van vragen

- Minimaal 8 vragen bij minimaal 5 antwoordcategorieën
- Minimaal 12 vragen bij minimaal 3 antwoordcategorieën

5.3 Open vragen

Omdat het goed invullen van een open vraag relatief veel moeite kost, kan een gemakzuchtige of frauduleuze respondent waarschijnlijk ook herkend worden aan de wijze waarop dergelijke vragen worden beantwoord. Een kwaadwillende respondent kan deze vragen ontwijken door het selecteren van “weet niet”, of het snel intypen van wat lukrake lettertekens. Om dergelijke responses te kunnen herkennen zouden de open antwoorden idealiter dienen te worden geanalyseerd op inhoud. Aangezien het onze eerste prioriteit was de mogelijkheden van een geautomatiseerd systeem te onderzoeken, is het niet beantwoorden van open vragen voor ons onderzoek gedefinieerd als ‘het niet beantwoorden van minimaal drie open vragen’.¹

¹ Dit is exclusief vragen die op basis van vragenlijstrouting niet worden gesteld in dezelfde vragenlijst.

5.4 Screeningsvragen

Om kans te maken op een maximale hoeveelheid incentives tegen een minimale inspanning verwachten we dat een frauduleuze respondent aan het begin van een vragenlijst relatief veel “ja” antwoordt op screeningsvragen, ze willen immers niet het risico lopen uit de screening te vallen. Aan het eind van een vragenlijst zal een frauduleuze respondent daarentegen relatief veel “nee” antwoorden, om tijdrovende vervolgvragen te voorkomen.

In dit onderzoek zijn alleen screeningsvragen aan het begin van de vragenlijst betrokken. Een respondent die bij een multiple response screeningsvraag alle antwoordcategorieën aanklikt is ten behoeve van ons onderzoek als mogelijk fraudeleus gekenmerkt.

5.5 Weet niet

Je zou kunnen verwachten dat een frauduleuze respondent relatief vaak “weet niet” aanvinkt, maar valide respondenten kunnen uiteraard ook net zo goed gebruik maken van deze categorie. In dit onderzoek is het gebruik van de “weet niet” categorie voorlopig buiten beschouwing gelaten.

5.6 Consistentiechecks en strikvragen

Een interessante methode om frauduleuze respondenten op te sporen is het onderling vergelijken van vragen en te controleren op onderlinge consistentie. Een nadeel van deze methode is dat per vragenlijst andere vragen vergeleken dienen te worden, wat een geautomatiseerde herkenning bemoeilijkt. In principe is het ook mogelijk om een gelijkblijvende set vragen te ontwikkelen die in elke vragenlijst meeloopt om frauduleuze respondenten op te sporen. Een bijkomende complicatie in panelonderzoek is echter dat een dergelijke set strikvragen regelmatig veranderd zou dienen te worden om irritatie en gewenning te voorkomen. Mede hierom is ten behoeve van dit onderzoek het gebruik van consistentiechecks en strikvragen vooralsnog buiten beschouwing gelaten.

5.7 Gele en rode kaarten

Voor ons onderzoek is gekozen voor de volgende bovengenoemde vier indicatoren: invulduur, straight lining, open vragen en screeningsvragen. Respondenten krijgen een “gele kaart” als ze op één of meer van deze variabelen scoren. In de volgende paragraaf worden deze resultaten nader besproken.

Een gele kaart hoeft niet noodzakelijkerwijs te duiden op fraude. Een respondent kan bijvoorbeeld werkelijk het antwoord op drie open vragen niet weten. In paragraaf 7 wordt hier nader op ingegaan. Hoe vaker een panellid een gele kaart krijgt, des te verdachter deze wordt. Bij veel gele kaarten zou een rode kaart verwijdering uit het panel kunnen betekenen. Hoewel

het interessant zou zijn om herhaaldgedrag in onze analyses mee te nemen was dat in het kader van dit artikel helaas niet mogelijk in verband met het beperkt aantal onderzoeken in de analyse.

6. Uitkomsten

6.1 Kwaliteitscriteria

In tabel 1 word het aantal respondenten weergegeven dat met elk van de vier kwaliteitscriteria is “gevlagd”. Opvallend is ten eerste dat voor elk van de indicatoren maar relatief weinig respondenten zijn geteld. De invulduur en de open antwoorden komen relatief het meest voor: een te korte invulduur is 907 maal geteld (1,5% van n=59.085) en weet niet bij drie of open antwoorden komt 912 keer voor (1,5%). Straight lining in twee of meer vraagbatterijen kwam maar in 323 cases voor (0,5%), terwijl het aanvinken van alle selectievragen maar heel weinig voorkwam: n=20 (0,0%).

Een heel onverwachte uitkomst van tabel 1 is dat er weinig of geen overlap tussen de indicatoren voorkomt. Wij hadden verwacht dat als er systematisch wordt gefraudeerd door bepaalde respondenten, dat zij op meerdere indicatoren tegelijkertijd zouden scoren. Wie bijvoorbeeld een vragenlijst extreem snel invult zal dan ook relatief veel straight linen of drie maal weet niet scoren op open antwoorden. Maar dit was niet het geval. De vraag is dan: kunnen we wel spreken van “rotte appels” als er geen bewijs is van consistent en systematisch dubieus antwoordgedrag? Het omgekeerde lijkt eerder het geval, respondenten die drie of meer open vragen open laten hebben een langere invulduur (7,96 sec per antwoord) dan andere respondenten (gem 5,00 sec; $t=28.95$, $p<0,001$).

Tabel 1: Aantal scores per indicator (n=59,085)

	Responsetijd	Straight lining	Open vragen	Selectievragen	Totaal
Response tijd	905	2	0	0	907
Straight lining	2	316	4	1	323
Open vragen	0	4	904	4	912
Selectievragen	0	1	4	15	20
Totaal	907	323	912	20	2162

6.2 Analyse naar achtergrondkenmerken van respondenten

Hoewel systematisch *satisficing* gedrag niet of nauwelijks voorkwam, waren we wel benieuwd naar de achtergrond kenmerken van respondenten die een of meerdere gele kaarten hadden gekregen. Worden bepaalde typen respondenten vaker gevraagd dan andere? Om dit na te gaan zijn van de 38.969 respondenten waar de 59.085 cases op zijn gebaseerd profielen gecalculleerd op een groot aantal achtergrondkenmerken.

Uit deze analyse bleek dat er nauwelijks sprake is van een samenhang tussen de scores op één van vier indicatoren en de achtergrondkenmerken. Er werden alleen kleine verschillen gevonden voor geslacht, samenstelling van het huishouden en beroep (zie tabel 2). Mannen werden iets meer gevraagd dan vrouwen, panelleden in oudere huishoudens meer dan personen uit andere soorten huishoudens en studenten minderen vaak dan andere beroepsgroepen. Verschillen met alle andere achtergrondkenmerken waren niet significant.

Tabel 2: Achtergrondkenmerken van gevraagde respondenten

	gele kaart		geen gele kaart	
	Abs.	%	Abs.	%
Mannen	501	47,6	15.744	41,5
Vrouwen	551	52,4	22.174	58,5
hoofdkostwinner > 34 jr.	471	35,7	11.167	29,9
Overige	850	64,3	26.236	70,1
Studenten	176	13,3	7.401	19,7
Overige	1147	86,7	30.246	80,3

6.3 Analyse naar kenmerken van de onderzoeken

Respondenten die op één of meer kenmerken van antwoordgedrag zijn gevraagd blijken in achtergrondkenmerken nauwelijks af te wijken van de overige leden van het Intomart GfK

online panel. Maar hoe zit dat als we een onderlinge vergelijking maken tussen de verschillende vragenlijsten? Bij vergelijking van het aandeel respondenten die op één of meer van de indicatoren voor antwoordgedrag zijn gevraagd tussen de 86 onderzoeken vallen een aantal interessante conclusies te trekken.

Bij slecht antwoordgedrag ten gevolge van een klein groepje kwaadwillende respondenten zou je een gelijke verdeling in het aantal gele kaarten per onderzoek verwachten. Deze verdeling blijkt echter zeer ongelijk. In tabel 3 is te zien dat in tweederde van de onderzoeken geen enkele respondent werd gevraagd, terwijl in 12 procent van de onderzoeken meer dan de helft van de respondenten een gele kaart kregen. Hier zou je uit kunnen concluderen dat de kwaliteit van het antwoordgedrag in online onderzoek eerder verbeterd kan worden door kritisch naar de onderzoeken te kijken dan door kritisch de panelleden te volgen.

Tabel 3: Aantal gele kaarten per onderzoek

gele kaarten	Aantal	%
geen	55	64%
weinig (0,1%-2%)	12	14%
gemiddeld (3%-10%)	3	3%
veel (11%-50%)	6	7%
zeer veel (50%+)	10	12%
Totaal	86	100%

Van de 86 onderzoeken waren er 19 waarbij 3% of meer van de steekproef een gele kaart kreeg. Bij tien van deze onderzoeken werd meer dan 50% van alle respondenten met een gele kaart gemarkeerd, één onderzoek veroorzaakte meer dan 50% van alle gele kaarten met betrekking tot de vragenlijstduur. Bij 9 van de 86 onderzoeken werd er gevraagd voor het niet beantwoorden van open vragen.

De gemiddelde invulduur neemt een belangrijke plaats in met totaal 42% van alle gele kaarten. In tabel 4 is per onderzoek het percentage gele kaarten ten gevolge van te korte invulduur uitgesplitst naar de totale duur van de vragenlijst. Bij relatief lange vragenlijsten van 8 minuten of meer is het percentage gele kaarten ten gevolge van een te korte invulduur relatief hoog, namelijk rond de 5%. Dit zou verklaard kunnen worden als respondenten als een vragenlijst langer duurt steeds sneller gaan invullen, omdat ze ongeduldig worden en een lange vragenlijst eerder verveelt. Het percentage *straight lining* en gele kaarten ten gevolge van open antwoorden blijkt ook toe te nemen naarmate vragenlijsten langer duren.

Het hoogste percentage gele kaarten ten gevolge van een te korte invulduur wordt echter gevonden bij heel korte vragenlijsten tot 1 minuut, waarbij pas een gele kaart is gegeven bij minimaal 30 gegeven antwoorden. Het is mogelijk dat deze gele kaarten niet veroorzaakt zijn door slecht antwoordgedrag. Bij korte vragenlijsten is het percentage *straight lining* en gele kaarten ten gevolge van open antwoorden juist relatief laag. De oorzaak van het hoge

percentage gele kaarten ten gevolge van een te korte invulduur kan liggen in afwijkingen van de juiste vaststelling van de vragenlijstduur, die ten behoeve van het meten van de kwaliteitsindicatoren voor het antwoordgedrag met name bij korte vragenlijsten exacter zou moeten geregistreerd. We komen hier in paragraaf 7 op terug.

Tabel 4: Verdachte responses naar invulduur (n=59.085, minimaal 30 responses in vragenlijst)

	Gele kaart	N
0-1 min	19.3%	3,690
1-2 minuten	0.9%	7,556
2-4 minuten	0.4%	13,063
4-8 minuten	2.0%	15,096
8-12 minuten	4.8%	8,525
12+ minuten	5.6%	11,155
Total	3.6%	59,085

Het percentage gele kaarten per onderzoek blijkt niet samen te hangen met de steekproefgrootte: onderzoeken met kleine en grote steekproeven verschillen niet in het aantal gele kaarten.

In tabel 5 is een eerste aanwijzing zichtbaar dat het aandeel dubieuze responses kan worden verklaard uit het onderzoeksdesign. Er worden duidelijk minder gele kaarten uitgedeeld bij onderzoek in opdracht van de overheid of non-profit organisaties, mogelijk omdat respondenten zich bij dergelijk onderzoek meer betrokken voelen. Bij opdracht in opdracht van adverteerders (marketing en consumenten onderzoek) en met name bij tracking onderzoek worden de meeste gele kaarten veroorzaakt door het niet invullen van open vragen. Dit kan mede veroorzaakt zijn omdat de open vragen hier veelal gaan om het noemen van merken en door respondenten misschien niet alleen saaier, maar ook als “moeilijker” wordt ervaren. Bij onderzoek in opdracht van media organisaties worden de meeste gele kaarten veroorzaakt door het te snel invullen van de vragenlijst.

Tabel 5: Gele kaarten naar soort opdrachtgever (n=59.085)

Project type	n	Gele kaarten	Invulduur	Straight lining	Open vragen	Screenings - vragen
Non profit/overhead	17.915	0,3%	0,0%	0,2%	0,0%	0,0%
Marketing-consumenten onderzoek	26.881	4,7%	0,6%	1,0%	3,0%	0,0%

Media	14.293	5,9%	5,1%	0,1%	0,6%	0,1%
Totaal	59.089	3,6%	1,5%	0,5%	1,5%	0,0%

Na afloop van het invullen van het onderzoek wordt in het kader van de kwaliteitsmonitoring van het Intomart GfK online panel een random deel van iedere steekproef gevraagd om door middel van een rapportcijfer een algemene waardering te geven voor de vragenlijst en deze te beoordelen op interesse, begrijpelijkheid, duur en de mate van herhaling. In tabel 6 worden deze evaluaties uitgesplitst naar soort gele kaart. Respondenten met een gele kaart zijn duidelijk veel negatiever over de vragenlijst, hun gemiddelde waardering is veel lager dan voor respondenten zonder gele kaart (M=6,51 versus M=7,09; t=5.18, p<0,01). Ook vindt de groep met een gele kaart de betreffende vragenlijst minder interessant en minder begrijpelijk. Tenslotte ervaart deze groep de vragenlijst als te lang met teveel herhaling. Vooral respondenten die open vragen niet invullen hebben een relatief negatieve mening over de vragenlijsten.

Tabel 6: Evaluatie vragenlijst naar soort gele kaart (n=7.498)

Soort gele kaart	Gem. waardering	Interessante vragenlijst*	Begrijpelijke vragenlijst *	Vragenlijst te lang*	Te veel herhaling in vragenlijst*
Gele kaart (n=266)	6,5	35,3%	53,8%	32,0%	50,8%
Invulduur (n=44)	6,9	38,6%	61,4%	20,5%	25,0%
Straight lining (n=57)	6,9	45,6%	50,9%	26,3%	38,6%
Open vragen (n=163)	6,3	30,7%	52,8%	36,8%	62,0%
Totaal (n=7,498)	7,1	49,6%	72,0%	17,6%	29,4%

* percentage *mee eens* of *zeer mee eens*

Deze resultaten suggereren dat respondenten onderzoekers kunnen helpen in het identificeren van problematische vragenlijsten. Als een onderzoek als te saai, te lang of te onbegrijpelijk wordt ervaren wordt dit door respondenten indien hiernaar gevraagd duidelijk aangegeven. We pleiten dan ook voor een breed gebruik van dergelijke evaluatievragen. Beter nog is het om deze vragen op te nemen in pilot onderzoek en deze maten serieus te nemen voordat de definitieve versie van een vragenlijst het veld in gaat.

Bij onderzoek naar frauduleuze respondenten en de kwaliteit van antwoordgedrag heeft het geen zin de oorzaak te zoeken bij de respondent. Door onze panelleden serieus te nemen en te luisteren naar hun inbreng kunnen we als onderzoekers veel leren om de vragenlijst relevanter en zinvoller te maken. Het zijn de respondenten die ons kunnen wijzen op mogelijk problematische vragenlijsten. De incentive voor de onderzoekers is duidelijk: we krijgen hiermee een betere kwaliteit data.

7. FLAGS in de praktijk

Na het in dit artikel beschreven vooronderzoek zijn de vier kwaliteitscriteria voor antwoordgedrag invulduur, *straight lining*, open antwoorden en selectievragen geïmplementeerd bij al het onderzoek in het Intomart GfK online panel als de FLAGS kwaliteitscontrole: “the GfK Survey and Panellist Quality Filter”. Een kwaadwillende respondent die zonder de vragen te lezen even snel een vragenlijst er doorheen “jast” loopt automatisch tegen de lamp en wordt uit de rapportage gehouden. De keuze om bij een gele kaart een mogelijk frauduleuze respondent uit de steekproef te verwijderen mag echter nooit een automatisme zijn. Het is immers altijd mogelijk dat het valide antwoorden betreffen. Een respondent kan bijvoorbeeld werkelijk dezelfde mening hebben op alle items in een vragenbatterij of geen antwoord weten op drie of meer open vragen. Bij elke vragenlijst dienen daarom onderzoekers te bepalen of de FLAGS kwaliteitsfilters relevant zijn.

Bij het in gebruik nemen van FLAGS zijn een aantal belangrijke hobbels genomen. De vier kwaliteitsindicatoren kunnen op al het online onderzoek automatisch worden toegepast. Uit het vooronderzoek bleek ondermeer dat de voor de kwaliteitsindicatoren noodzakelijke parameters exact dienen te worden vastgesteld. Een voorbeeld hiervan was de vragenlijstduur. Deze werd aanvankelijk op basis van de inschatting van de onderzoeker van te voren vastgesteld, waarbij onder meer met effecten als routing rekening werd gehouden. Uit het vooronderzoek bleek dat ook onderzoekers niet onfeilbaar zijn: het bleek lastig om de duur van een vragenlijst correct in te schatten, waardoor met name bij korte vragenlijsten de kwaliteitsindicator op invulduur te scherp werd ingesteld. Inmiddels wordt deze parameter tijdens het veldwerk in real time vastgesteld.

We besluiten dit artikel met een praktijkvoorbeeld. Eén van de 86 besproken onderzoeken betrof een vragenlijst van gemiddeld 12 minuten over een veel gekocht consumentenproduct onder 374 gebruikers van de betreffende productcategorie. De vragenlijst betrof (afhankelijk van de routing) ongeveer 10 stellingblokken over onder meer bekendheid met producten, aankoopintentie, mening over een aantal verschillende merken en tot slot algemene stellingen over prijs en kwaliteit van het betreffende consumentenproduct. In de evaluatie kreeg het onderzoek een rapportcijfer 6,9 (ten opzichte van een 7,1 voor de gemiddelde vragenlijst). In tabel 7 worden de overige evaluatiematen voor de productvragenlijst weergegeven.

Tabel 7: Evaluatie van de productvragenlijst

	Product- vragenlijst	Overige vragenlijsten
Interessante vragenlijst	38%	42%
Begrijpelijke vragenlijst	63%	74%
Vragenlijst te lang	32%	17%
Te veel herhaling in vragenlijst	46%	24%

Het bleek dat de productvragenlijst goed te begrijpen is en niet veel minder interessant dan de meeste andere onderzoeken. De vragenlijst werd echter ook relatief als te lang en met te veel herhaling ervaren. Uit de open antwoorden wat men van het onderzoek vond kwamen stellingen als “*Heel saai. Sommige dingen zijn moeilijk te onderscheiden, vooral als je niet echt geïnteresseerd bent in het onderwerp*” en “*Ik heb wel eens leukere vragenlijsten gezien!*”, maar ook “*Ik vind het leuk om in het panel te zitten en ben iedere keer weer nieuwsgierig naar wat jullie vragen*” en “*Ik vond de vragenlijst helder en niet te lang. Ga vooral zo door*”.

Geen geweldig goede evaluatie, maar ook geen reden om te twifelen aan de kwaliteit van het onderzoek. Maar als we voor dit onderzoek naar de FLAGS criteria kijken, ontstaat een heel ander beeld (zie tabel 8).

Tabel 8: FLAGS criteria bij evaluatie van een productvragenlijst

	Product- vragenlijst	Overige vragenlijsten
Gele kaart	34%	3,6%
Invulduur	0%	1,5%
Straight lining	34%	0,5%
Open vragen	1,5%	1,5%

De productvragenlijst valt bij controle door middel van FLAGS direct op door een wel zeer grote mate van straight lining bij maar liefst 126 van de 374 respondenten. Bij het kritisch lezen van de vragenlijst wordt ook wel duidelijk waarom zoveel respondenten hun geduld verloren bij het invullen van vragenlijst. Het was ook voor ons geen pretje om ruim 10 minuten lang voor 4 merken die sterk op elkaar lijken en die je niet elk even goed kent fikse vraagbatterijen te moeten invullen met elk een stuk of 12 items. Dit soort saaie vragenlijsten kunnen ontstaan door een heel legitieme behoefte van een opdrachtgever en een onderzoeker die de tijd niet nam om de vragenlijst aan te passen of aan te vullen met aantrekkelijker vragen. De vragenlijst viel door de mand bij één derde van de respondenten, die besloten geen moeite te steken in zo'n saaie vragenlijst en misschien te beleefd bleven bij de evaluatie van het onderzoek. Dat dit onderzoek toch tegen de lamp liep is te danken aan de automatische FLAGS controle. Bij herhaling van het onderzoek is de vragenlijst in overleg met de opdrachtgever zodanig aangepast dat de invul kwaliteit tot een acceptabel niveau kon worden verbeterd.

Bij controle achteraf bleek overigens dat het eventueel verwijderen van de groep van 126 respondenten met straight lining geen statistisch significant effect zou hebben gehad op de gerapporteerde gemiddelden voor de kernvariabelen merkbekendheid, koopintentie en

beoordeling van de prijs/kwaliteit verhouding van de twee belangrijkste merken uit het onderzoek. Dit is een aanwijzing dat straight lining en andere vormen van ongewenst invulgedrag random over de strata zijn verdeeld. Dat deze ruis geen effect heeft op de gerapporteerde gemiddelden mag uiteraard geen argument zijn om dergelijke controles achterwege te laten.

8. Conclusies, discussie en aanbevelingen

Hoewel qua analyses nog niet alles uit de kast is gehaald, kunnen we in ieder geval één belangrijke conclusie trekken: *satisficing* komt met enige regelmaat voor, maar dat zegt meestal méér over de aard van de vragenlijst dan over het gedrag van de respondent of het panellid. Voor ons betekent dit dat de belangrijkste factor bij het bewaken van datakwaliteit de onderzoeker zélf is. Als er al iets verdacht is, is het waarschijnlijk niet zozeer de respondent, maar is de kans groot dat vragenlijsten antwoorden *uitlokken* die als verdacht worden beschouwd. Dit betekent voor de onderzoekers dan ook dat grote voorzichtigheid moet worden betracht bij activiteiten binnen de branche die als doel hebben om van dubieuze respondenten af te komen. In principe is dit een positieve conclusie. We zijn niet zo zeer afhankelijk van de grillen van respondenten, maar hebben voor het overgrote deel de kwaliteit van onze data zelf onder controle.

Het feit dat nauwelijks respondenten zijn gevraagd bij méér dan één van de indicatoren per vragenlijst geeft aan dat de panelleden van het Intomart GfK Online Panel veel sterker *optimizing* dan *satisficing* laten zien. Als dit niet het geval was, waren we namelijk veel meer respondenten tegengekomen die zich uit de voeten maken met bijvoorbeeld zowel straight lining én het niet beantwoorden van open vragen. Maar dergelijke combinaties zijn niet of nauwelijks gevonden. Hoewel we tevreden kunnen zijn met dit resultaat, realiseren wij ons tevens dat het aantal projecten binnen dit onderzoek relatief klein is. De datakwaliteit zal dan ook gecontroleerd moeten blijven worden op zowel het niveau van de respondent als van de vragenlijst.

Er blijft dus werk aan de winkel, de kwaliteit van antwoordgedrag verdient continu aandacht. Ten eerste zullen we blijven onderzoeken en experimenteren met manieren om *optimizing* gedrag te stimuleren. Het in detail bestuderen naar het waarom respondenten naar *satisficing* gedrag neigen zal aanknopingspunten bieden voor zowel verder onderzoek als voor manieren om de datakwaliteit te bevorderen. Tegelijkertijd blijven we op zoek naar heldere en vooral objectievere criteria om verdacht gedrag op te sporen.

Vooruit kijkend hopen we dat de kwaliteit van zowel vragenlijsten als panelleden branchebreed hoog op de agenda worden gezet. Hoewel deze studie niet direct kan worden gegeneraliseerd naar andere online panels, zijn de auteurs er van overtuigd dat onderzoekers, panelmanagers en panelgebruikers hun verantwoordelijkheid serieus moeten nemen om datakwaliteit te kunnen blijven waarborgen. Met in het achterhoofd de continue dalende

bereidheid om deel te nemen aan vragenlijstonderzoek, de stijgende vragenlijstdruk en paneloverlap, kunnen we het ons simpelweg niet permitteren om systematisch panelleden uit te sluiten van vragenlijstonderzoek vóórdat we als onderzoeker er alles aan gedaan hebben om het vragenlijstproces te optimaliseren. We zullen daarbij nog meer door de bril van de respondent moeten kijken. Aangezien we als branche voor hetzelfde probleem staan, vinden wij het belangrijk om naar consensus te streven wat betreft hoe datakwaliteit (1) te meten, (2) te verbeteren en (3) te handhaven.

Vóór de activiteiten die in dit paper zijn beschreven heeft Intomart GfK kwesties omtrent datakwaliteit ad hoc benaderd. Deze studie beschrijft onze eerste ervaringen met het op systematische en organisatiebrede wijze evalueren. We hebben geëxperimenteerd met vragenlijstonafhankelijke indicatoren; dat wil zeggen elke indicator is voor elk onderzoek op dezelfde manier geoperationaliseerd. Hoewel helder en pragmatisch, is dit gezien de variabiliteit tussen het type projecten mogelijk niet de optimale manier om datakwaliteit in kaart te brengen. Het lijkt ons echter praktisch gezien onmogelijk om elk project individueel en ad hoc te evalueren. Om de datakwaliteit van alle surveys binnen de organisatie te monitoren is een zekere mate van efficiency onvermijdelijk. Daarom zullen we er naar streven de indicatoren en criteria verder aan te scherpen zodanig dat geen handmatige operationalisaties, analyses, interpretaties en beslissingen voor elk individueel project zijn vereist.

Ter afsluiting willen we graag een aantal aanbevelingen doen, met als streven de datakwaliteit bij online onderzoek branchebreed te verbeteren:

- Beperk de lengte van de vragenlijst; wees bij online onderzoek extra alert bij vragenlijsten van langer dan 10 minuten.
- Wees extra alert bij tracking onderzoek; dit type onderzoek is over het algemeen niet het meest respondentvriendelijke.
- Voer standaard een pilot studie uit om de kwaliteit van de vragenlijst te verbeteren. Panelleden die hiervoor expliciet worden uitgenodigd zien zichzelf als ambassadeurs van het panel en dat garandeert serieuze en bruikbare feedback. Dit geeft de onderzoeker ook de mogelijkheid om binnen afzienbare tijd en met een relatief kleine steekproef een dergelijke pilot uit te voeren.
- Maak opdrachtgevers bewust van kwesties rondom datakwaliteit (korte termijn) en duurzaamheid van de panelbranche (lange termijn). Kom tot overeenstemming, bij voorkeur vóórdat het onderzoek begint, over de do's en don'ts van vragenlijsten. Bij lange saaie vragenlijsten van 45 minuten over productgebruik en attitudes ten aanzien van merken dreigt onherroepelijk het gevaar van slechte responsekwaliteit.
- Zet datakwaliteit (korte termijn) en duurzaamheid van de panelbranche (lange termijn) op de agenda binnen de branche. Zoals eerder gesteld, dit is een branchebreed probleem dat ook om een branchebrede oplossing vraagt. De auteurs hopen met deze studie in ieder geval een bijdrage te leveren en zullen dit ook de komende jaren blijven doen.

Literatuur

- Bowman. J. (2007). Hands up. who's on an online panel? Research World April 2007. Esomar.
- Callegaro. M., Yang. Y., Bholra. D.S. & Dillman. D.A. (2005). Response Latency as an Indicator of Optimizing. A Study Comparing Job Applicants and Job Incumbents' Response Time on a Web Survey. Working Paper series of the Program in Survey Research and Methodology University of Nebraska. Lincoln & Gallup Research Center (<http://sram.unl.edu>).
- Comley. P. (2006) The Games We Play. A psychoanalysis of the relationship between panel owners and panel participants. Panel Research 2006. ESOMAR World Research Conference.
- Downes-Le Guin. T. (2005). Satisficing Behaviour in Online Panel members. White paper Doxus.
- Knapton. K. & Garlick. R. (2006). Catch Me If You Can! How to Keep Cheaters Out of Your Next Online Survey. Presentation at The Market Research Event. October 23 2006. Los Angeles.
- Krosnick. J. A. (1999). Survey Research. Annual Review of Psychology. 50(1). 537-567.
- Krosnick. J. A. & Alwin. D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. Public Opinion Quarterly. 51(2). 201-219.
- Meurs, A. van; Ossenbruggen, R. van & Nekkers, L. (2007) Do rotten apples spoil the whole barrel? Exploring quality issues in panel data. In: ESOMAR Panel Research 2007 Conference (CD-ROM). Amsterdam: ESOMAR.
- Smith. R. & Hofma Brown. H. (2006). Panel and Data Quality. Comparing Metrics and Assessing Claims. Paper presented at ESOMAR Panel Research 2006. Barcelona.
- Smith. R. & Hofma Brown. H. (2005). Assessing the Quality of Data from Online Panels: Moving Forward with Confidence. White paper. Harris Interactive.
- Sudman, S., Bradburn, N., & Schwarz, N. (1996). Thinking about answers: The application of cognitive processes to survey methodology. San Francisco, CA: Jossey-Bass.
- Tourangeau, R., Rips. L.J. & Rasinski, K. (2000). The psychology of survey response. Cambridge University Press.
- Willems. P., Ossenbruggen. R. van. Vonk. T. (2006). The Effects of Panel Recruitment and Management on Research Results. A study across 19 online panels. Paper presented at ESOMAR Panel Research 2006. Barcelona.

Curriculum vitae

Lex van Meurs (lex.van.meurs@gfk.com)

Lex van Meurs werkt bij Intomart GfK aan de methodologie en ontwikkeling van media-onderzoek en is als research manager verantwoordelijk voor het online Waarderingspanel. Hij is tevens parttime Universitair Docent aan de Afdeling Communicatiewetenschap van de

Universiteit van Amsterdam. Daar studeerde hij methoden en technieken van politicologisch onderzoek en communicatiewetenschap (cum laude) en promoveerde met het proefschrift "Switching during commercial breaks".

Lex van Meurs won drie maal de Hans du Chatinier Prijs van de MWG (thans AMMA) voor het beste, meest innovatieve en meest toepasbare media onderzoek met "Aandacht voor Televisie" (1995), het SPOT-onderzoek naar schakelgedrag tijdens reclameblokken (1997), en het Bereiksonderzoek buitenreclame (2000).

Hij publiceerde ook over onder meer de evaluatie van meetinstrumenten, segmentatie-analyses, het voorspellen van kijkdichtheid, verkiezingscampagnes op televisie, het Kijkonderzoek en effectiviteit van buitenreclameposters.

Lex van Meurs houdt van diepzeeduiken, wandelen, yoga en opera.

Robert van Ossenbruggen. ProCression (robert@procreSSION.nl)

Robert van Ossenbruggen studeerde psychologie en marketing en adviseert sinds 2004 diverse marktonderzoekbureaus en opdrachtgevers onder de vlag van ProCression. Hij initieerde in samenwerking met vakgenoten Pieter Willems en Ted Vonk het NOPVO, oftewel het Nederlands Online Panel VergelijkingsOnderzoek.

Robert van Ossenbruggen is voorzitter van de MOA vakgroep Methoden en Technieken en schrijft regelmatig artikelen over uiteenlopende methodologische onderwerpen voor MOA en ESOMAR. In 2005 won Van Ossenbruggen MOA's Master Talent Award.

Liesbeth Nekkers, Senior Media Researcher bij Intomart GfK (liesbeth.nekkers@gfk.com)

Liesbeth Nekkers werkt sinds 2000 bij Intomart GfK. Op dit moment is zij verantwoordelijk voor het continue kijkonderzoek dat wordt uitgevoerd in opdracht van de Stichting KijkOnderzoek. Eerder was zij binnen Intomart GfK betrokken bij de uitvoering en aansturing van Advertising Tracking onderzoek, het printbereiksonderzoek, de formatie van het Intomart GfK online panel en was zij werkzaam als adviseur voor alle internetgerelateerde projecten binnen Intomart GfK.

Liesbeth Nekkers studeerde Psychologie aan de Universiteit van Amsterdam, met als afstudeerrichting Psychologische methoden van onderzoek (1997). Daarna was zij werkzaam als wetenschappelijk onderzoeker bij het Paedologisch Instituut in Duivendrecht en als Media Onderzoeker bij Blauw Research in Rotterdam.

English summary: Do rotten apples spoil the whole barrel? Checking the quality of responses in the Intomart GfK online panel

To improve on the quality of online research Intomart GfK introduced the FLAGS system in 2007 that checks on fraudulent responses based on response time, straight lining, and answers to open questions and screening questions.

In this article the results of 86 online research projects from the Intomart GfK online panel are discussed. Based on the FLAGS definitions we only found 4% dubious responses. We determined that there is very little of a systematic nature to be found in the demographic profile of panel members that are flagged.

FLAGS also detects questionnaires that invoke dubious responses. This proves to be especially questionnaires that were evaluated as less interesting, too long or boring. The most effective measure to prevent fraudulent responses is the quality of the questionnaires.

FLAGS automatically detects fraudulent respondents and offers checks to work together with clients to raise the quality of responses by improving the questionnaires.

Keywords: keywords: online panels, panel fraud, data integrity, data quality, questionnaire design