

# **Kwaliteitsevaluatie van surveys**

## **Een toepassing op de surveys naar culturele verschuivingen in Vlaanderen**

*Geert Loosveldt en Ann Carton  
Departement Sociologie, K.U.Leuven*

### *1. Inleiding*

Na de gemeenteraadsverkiezingen van oktober 2000 was één van de commentaren en conclusies dat de peilingen naar kiesintenties terug gefaald hadden. De reeds kwalijke reputatie van dergelijke peilingen werd daarmee bevestigd en versterkt. De evaluatie van de kwaliteit van deze peilingen gebeurt door de voorspelde kiesintenties te vergelijken met de uiteindelijke uitslag van de verkiezingen. Het gaat om een eenvoudige outputevaluatie waarbij externe data gebruikt worden om de peilingsdata te valideren. Daarbij maakt men de terechte veronderstelling dat wanneer men naar kiesintenties peilt de resultaten van de peiling op toevalsfouten na gelijk moeten zijn aan de verkiezingsuitslag.

Een peiling naar kiesintenties kan beschouwd worden als een bijzonder inhoudelijke invulling van een survey-onderzoek waarbij een representatieve steekproef van respondenten bevraagd wordt aan de hand van een gestructureerde vragenlijst over hun toekomstig stemgedrag en hun houdingen en opinies in verband met politiek. Meestal zijn de werkelijke populatieverdelingen van de kenmerken die men door middel van een survey wenst te meten niet beschikbaar en kan een eenvoudige outputevaluatie, waarbij de resultaten van het survey getoetst worden aan externe data, niet gemaakt worden. De vraag kan dan ook gesteld worden hoe de kwaliteit van surveys kan geëvalueerd worden wanneer de werkelijke populatieverdelingen niet gekend zijn. Daarbij dient onmiddellijk opgemerkt te worden dat een outputevaluatie van een survey slechts een zeer beperkte evaluatie is. Bij een dergelijke evaluatie wordt het proces van de dataverzameling en data-analyse immers buiten beschouwing gelaten. Naast een outputevaluatie is er dus ook een procesevaluatie van een survey mogelijk.

Bij de duiding waarom peilingen naar kiesintenties falen gebeurt er trouwens impliciet een procesevaluatie. Men verwijst naar bepaalde aspecten van de dataverzameling (o.a. representativiteit van de steekproef, non-respons en antwoordgedrag van de respondenten) als vermoedelijke oorzaken van het falen van peilingen naar kiesintenties. Daarbij maakt men de veronderstelling dat het optimaal verlopen van het proces van dataverzameling een noodzakelijke voorwaarde is voor het bekomen van kwalitatief degelijke data. Dit illustreert dat een procesevaluatie van een survey niet onbelangrijk is bij de kwaliteitsevaluatie van surveys.

Zowel een procesevaluatie als een outputevaluatie van een survey kunnen geconcretiseerd worden door gebruik te maken van het concept 'Integrale kwaliteitszorg'. In het algemeen bestaat integrale kwaliteitszorg uit een geheel van activiteiten die werknemers overtuigt om permanente verbeteringen aan te brengen in hun werkproces, die hen motiveert om problemen te identificeren en die hen in staat stelt om oplossingen te ontwikkelen en te implementeren. Men stelt daarbij dat alle onderdelen van het productieproces afzonderlijk en in onderlinge samenhang direct of indirect bijdragen aan de kwaliteit van het eindproduct. Kwaliteit is daarbij een kwestie van gedeelde verantwoordelijkheid (Carton, Loosveldt & Billiet 1998; Carton, 1999). Deze algemene omschrijving van een kwaliteitsbenadering kan toegepast worden bij het uitvoeren van een proces- en outputevaluatie van een survey-onderzoek. We beperken ons in wat volgt tot de kwaliteitsevaluatie van surveys waarbij de dataverzameling gebeurt door middel van face-to-face interviews. De aandachtspunten die bij de 'integrale kwaliteitszorg' aan bod komen worden daarbij als leidraad gebruikt. Alle activiteiten tijdens de fase van de vragenlijstconstructie worden buiten beschouwing gelaten. Zowel de procesevaluatie als de output evaluatie zijn gerelateerd aan twee belangrijke componenten van de taak van de interviewer tijdens de dataverzameling. Die componenten zijn het contacteren en het verkrijgen van medewerking van de respondent enerzijds en het interviewen in enge zin anderzijds. Bij dit laatste gaat het om het stellen van vragen, het geven van instructies, doorvragen ...en het registreren van de antwoorden van de respondent. Het contacteren en het verkrijgen van medewerking resulteert in de 'gerealiseerde steekproef'. De geregistreerde antwoorden zijn de output van het interviewen in enge zin.

De procesevaluatie die gebeurt vanuit het concept van integrale kwaliteitszorg zal nagaan in welke mate interviewers de voorgeschreven regels en procedures volgen bij de uitvoering van de twee zonet vermelde componenten van hun taak. Daarbij zal ook worden nagegaan hoe hij/zij op zijn taak wordt voorbereid en welke feedback hij/zij krijgt bij de uitvoering ervan. Daarmee komen we bij de rol van de onderzoeker die ervoor moet zorgen dat de interviewers degelijk zijn voorbereid, beschikken over een actuele adressenlijst van de respondenten en gebruik kunnen maken van een degelijke vragenlijst. De onderzoeker is met andere woorden verantwoordelijk voor de input die de interviewer gebruikt bij het uitvoeren van zijn taak. Het is de combinatie van de kwaliteit van de input en de kwaliteit van de taakuitvoering die de kwaliteit van de output bepaalt.

Bij een outputevaluatie staat de kwaliteitsvraag centraal: in welke mate zijn de 'gerealiseerde steekproef' en de geregistreerde antwoorden vrij van fouten. De totale fout bij een survey onderzoek kan gedefinieerd worden als het verschil tussen een werkelijke populatiewaarde en een schatting ervan op basis van een steekproef. Er kunnen verschillende bronnen van fouten worden onderscheiden (Groves, 1989). Bij de gerealiseerde steekproef is de evaluatie van non-respons fouten een bijzonder belangrijk aspect van kwaliteitsbeoordeling. De non-respons fout is het gevolg van het feit dat de eenheden van de steekproef waarbij een non-respons (weigeren, niet te contacteren, ...) wordt gerealiseerd verschillend zouden hebben geantwoord dan diegene die hebben meegewerkt aan het survey interview. Bij de geregistreerde antwoorden dient een evaluatie van meetfouten te gebeuren. Meetfouten zijn inaccurate antwoorden die voortvloeien uit slecht geformuleerde vragen, inadequaat interviewergedrag en bepaalde aspecten van het respondentengedrag. Non-response fouten en meetfouten worden onder andere bepaald door de wijze waarop de interviewer zijn taak uitvoert. Bij een outputevaluatie die zich richt op de verschillende soorten fouten die kunnen

optreden, gebeurt er dus indirect een evaluatie van het interviewergedrag. Het is indirect omdat het gaat om een evaluatie op basis van de verzamelde data.

In het kader van integrale kwaliteitszorg worden de resultaten van een procesevaluatie en een outputevaluatie gebruikt om het onderzoeksproces verder te optimaliseren. Wel dient men er zich van bewust te zijn dat de kwaliteitszorg in de context van survey-onderzoek beperkingen kent.

Niet alle factoren die de kwaliteit bepalen van een survey-onderzoek in het algemeen of een opiniepeiling in het bijzonder kunnen door een onderzoeker gestuurd of gecontroleerd worden. Dergelijk onderzoek gebeurt nu eenmaal in een maatschappelijke context waarin een bepaald klimaat heerst dat bevorderlijk of hinderlijk is voor het al of niet meewerken aan dergelijk onderzoek. Ook de wijze waarop de respondent reageert op het verzoek om mee te werken en de mate waarin hij gemotiveerd is om tijdens een interview zijn rol van respondent adequaat te vervullen, kunnen door de onderzoeker slechts in zeer beperkte mate worden 'gemanipuleerd' in functie van een optimaal verloop van het onderzoeksproces.

In wat volgt gebeurt er in eerste instantie een outputevaluatie en vervolgens een procesevaluatie van de geregistreerde antwoorden en de gerealiseerde steekproef van een survey. In het onderstaand schema wordt alles nog eens samengevat en wordt ook duidelijk welke onderdelen van de kwaliteitsevaluatie zullen worden toegepast en uitgewerkt in het kader van het survey naar de culturele verschuivingen in Vlaanderen 2000.

*Schema: Kwaliteitsevaluatie van het survey naar culturele verschuivingen in Vlaanderen.*

Taken van de interviewer		
	Contacteren en verkrijgen van medewerking	Interviewen in enge zin
Resultaat uitvoering	Gerealiseerde steekproef	Geregistreerde antwoorden
Procesevaluatie	Evaluatie contact- en vervangingsprocedure	Evaluatie training en opvolging. Evaluatie interviewergedrag
Outputevaluatie	Evaluatie non-respons fout	Evaluatie meetfouten

## 2. Outputevaluatie van de gerealiseerde steekproef

Een bijzonder belangrijk onderdeel van een survey onderzoek is de trekking van een representatieve steekproef. Een precieze populatieomschrijving, een volledig en adequaat steekproefkader en het correct toepassen van een steekproefprocedure zijn daarbij noodzakelijk. De steekproeffout en de dekkingfout zijn fouten die verbonden zijn met de technische (o.a. keuze voor het type steekproef) en concrete (o.a. steekproefkader) uitwerking van een steekproef. De steekproeffout is de fout die ontstaat door het feit dat er gewerkt wordt met een

steekproef waardoor de schatting ‘slechts’ op toevalsfouten na gelijk is aan de populatiewaarde. Het gaat om de precisie van een schatting die in belangrijke mate wordt bepaald door de steekproefomvang en die kan worden uitgedrukt door middel van een betrouwbaarheidsinterval. Ook bij een ‘perfecte’ steekproef hebben we te maken met een steekproeffout. Een dekkingsfout ontstaat wanneer niet alle eenheden van de populatie waarover men uitspraken wenst te doen, een gekende berekenbare kans (verschillend van nul) hebben om in de steekproef te worden opgenomen.

De gerealiseerde steekproef bestaat uit alle onderzoekselementen van de oorspronkelijk getrokken steekproef waarbij er tijdens de dataverzameling ook daadwerkelijk een interview werd afgenomen. Omwille van een mogelijke non-respons fout is de gerealiseerde steekproef niet noodzakelijk gelijk aan de oorspronkelijk getrokken steekproef. In de inleiding werd de non-respons fout reeds gedefinieerd als het verschil tussen diegene die hebben meegewerkt aan het survey interview en zij die niet meewerkten. Dit verschil impliceert dat de non-respons niet toevallig gebeurde. Het verschil tussen respondenten en zij die niet meewerkten voor een inhoudelijk relevant kenmerk en de omvang van de non-respons bepalen de non-respons fout bij de meting van dit kenmerk. Indien zij die niet meewerken niet significant verschillen van de respondenten dan treedt er geen non-respons fout op ongeacht het non-respons percentage. Een hoog percentage non-respons betekent dus niet noodzakelijk een grote non-respons fout. Het is duidelijk dat zowel het (non)respons percentage als het verschil tussen respons en (non)respons bij een outputevaluatie van de gerealiseerde steekproef aan bod dient te komen.

### *2.1 Evaluatie van het (non)respons percentage*

Binnen het methodologisch onderzoek naar de kwaliteit van survey data is er een groeiende aandacht voor non-respons. Uit een internationaal vergelijkend onderzoek blijkt dat er in een aantal Europese landen een dalende trend is van het percentage respondenten dat wil meewerken aan een survey-interview. Daarbij worden er verschillen vastgesteld tussen de landen. Deze verschillen worden onder andere bepaald door algemene kenmerken van het survey (o.a. onderzoeksdesign, waarnemingseenheid, gebruik van proxy respondenten,...) en de wijze waarop het veldwerk in concreto georganiseerd wordt (contactstrategie, aantal contacten, beloningssysteem voor de interviewers,...). Uit dit internationaal vergelijkend onderzoek blijkt ook dat er geen aanwijzingen zijn van een groeiende groep van harde weigeraars maar dat er wel grotere inspanningen (bv. meer contactpogingen,...) moeten gedaan worden om het responspercentage op eenzelfde peil te houden (De Heer, 1999; de Leeuw & De Heer, 1999).

Bij het evalueren en het eventueel vergelijken van (non)responspercentage is het belangrijk om bij de berekening aan te geven welke steekproefelementen er mee in rekening worden gebracht. Bij de berekening van het meest algemene respons percentage wordt er best gewerkt met alle steekproefelementen bij wie een interview had kunnen worden afgenomen (AAPOR, 2000).

In tabel 1 worden de (non)responsprofielen weergegeven van de vijf APS surveys die werden georganiseerd in de periode 1996-2000. Bij de berekening van de (non)respons percentages werd de categorie ‘niet inzetbare adressen’ buiten beschouwing gelaten. Deze categorie omvat de reden ‘overleden’, ‘verhuisd’, ‘taalbarrière’, en ‘niet in staat om deel te nemen wegens ziekte, handicap of dementie’. We beschouwen deze niet inzetbare adressen

als steekproefelementen bij wie er in feite geen interview kan worden afgenomen. In de vier surveys waar deze categorie kan worden onderscheiden gaat het telkens om ongeveer 7% van het totaal aantal gebruikte adressen. Daar het in feite gaat om mankementen van het steekproefkader kan die 7% beschouwd worden als een schatting van de dekkingsfout. De categorie 'andere non-respons' die wel in de tabel is opgenomen omvat de reden 'thuis maar echt niet in de gelegenheid' en een niet nader bepaalde reden 'ander'.

Verder wordt er een onderscheid gemaakt tussen de gebruikelijke categorieën 'afgewerkt interview', 'weigeringen' en 'niet te contacteren'. We beschouwen de frequentieverdeling over de vier onderscheiden categorieën als het (non)responsprofiel. Het totaal per jaar dat in tabel 1 wordt vermeld, heeft dus betrekking op het aantal steekproefelementen dat gebruikt werd en waarbij een interview had kunnen worden afgenomen.

Tabel 1: Overzicht van (non)responsprofielen in de APS surveys 1996-2000

	1996		1997		1998		1999		2000	
Vooropgesteld aantal Interviews	1500		1500		1510		1510		1500	
Afgewerkte interviews	1577	71.3	1500	60.2	1491	49.3	1376	45.7	1345	64.5
Weigeringen	425	19.2	497	19.9	1199	39.7	1008	33.5	453	21.7
Niet te contacteren	187	8.5	442	17.7	333	11.0	544	18.1	224	10.7
Andere nonrespons	22	1.0	53	2.1			83	2.7	64	3.1
Totaal	2211		2492		3023		3011		2086	

Behalve in 1998 stond het marktonderzoeksbureau Dimarso/NID in voor de uitvoering van het veldwerk, de codering en de eerste gegevensverwerking. In 1998 werd beroep gedaan op het marktonderzoeksbureau INRA/Marketing Unit. De selectie van de bureaus gebeurde op basis van een bestek en binnen het kader van de wet op de overheidsopdrachten.

Gegeven dat het vooropgesteld aantal te realiseren interviews contractueel is vastgelegd, is het merkwaardig dat dit aantal niet steeds gerealiseerd werd en dat het absoluut aantal afgewerkte interviews telkens daalde. Bij het eerste survey in 1996 werden er meer interviews afgenomen dan contractueel was vastgelegd. In 1999 en 2000 werden respectievelijk nog slechts 91.1% en 89.7% van het vooropgestelde aantal interviews gerealiseerd. Verder vertonen de responsprofielen van de vijf surveys grote schommelingen en zijn er tijdens die periode van vijf jaar geen duidelijke tendensen te onderkennen. Zo kan er niet gesteld worden dat het responspercentage (= percentage afgewerkte interviews) telkens daalt en het percentage 'weigeringen' en het percentage 'niet te contacteren' telkens stijgt. Een dergelijk patroon zou dan kunnen 'verklaard' worden door te stellen dat het steeds moeilijker is om mensen te contacteren en te overtuigen om mee te werken aan een interview. Deze verklaring stelt de respondent verantwoordelijk voor het (non)responsprofiel. Een dergelijke verklaring is echter niet afdoend voor de plotse stijging in 97 van het percentage 'niet te contacteren'. Door die plotse stijging daalt het responspercentage met meer dan 10%. Ook de sterke stijging van het percentage weigeringen in 1998 en de daling ervan in 1999 en 2000

passen niet in de verklaring. Het (non)responsprofiel van het survey van 2000 past trouwens helemaal niet in deze verklaring. De conclusie lijkt dan ook te zijn dat de impact van het gedrag van de respondenten bij de verklaring voor de schommelingen in de responsprofielen wellicht beperkt is. De verklaring voor de schommelingen dient zeker ook gezocht te worden in de wijze waarop het veldwerk van elk survey in concreto wordt georganiseerd. De contactprocedure, de wijze waarop interviewers over vervangadressen kunnen beschikken en de opvolging van de interviewers zijn daarbij belangrijke aspecten. Het is duidelijk dat het hier gaat om elementen waarop de procesevaluatie zich richt en waarvoor de onderzoeker verantwoordelijk is. De onderzoeker is met andere woorden mede verantwoordelijk voor het gerealiseerde responsprofiel. Tijdens de procesevaluatie van de gerealiseerde steekproef wordt dit verder uitgewerkt.

## *2.2 Evaluatie van het verschil tussen respons en non-respons*

Eerder werd er reeds op gewezen dat de responsfout niet alleen bepaald wordt door het percentage (non)respons maar ook door het verschil tussen respondenten en zij die niet meewerkten. Een centrale vraag bij de outputevaluatie is dan ook of zij die hebben meegewerkt aan het interview op inhoudelijk relevante kenmerken verschillen van zij die niet hebben meegewerkt. Het is niet evident om dit verschil te evalueren. Meestal is er geen informatie over die inhoudelijk relevante kenmerken beschikbaar voor diegene die niet meewerkten aan het survey. Toch zijn er een aantal mogelijkheden dit na te gaan. De meest voor de hand liggende werkwijze is de steekproef te vergelijken met de populatie waaruit die steekproef is getrokken.

### *2.2.1 Vergelijking tussen de gerealiseerde steekproef en de populatie*

Net zoals bij de reeds eerder vermelde outputevaluatie van peilingen naar kiesintenties kan de steekproefinformatie vergeleken worden met eventueel beschikbare populatie informatie. Dit gebeurt doorgaans met de beschikbare populatieverdeling van geslacht en leeftijd. Wanneer we de resultaten van de enquête naar de beroepsbevolking beschouwen als populatie informatie dan kunnen we ook voor opleidingsniveau een vergelijking maken. Dit gebeurt in tabel 2 waarin de verschillen zijn weergegeven tussen de steekproefverdeling van het hoogst behaalde diploma en de populatieverdeling van dit kenmerk. Een negatief teken van het verschil wijst dus op een ondervertegenwoordiging en een positief teken op een oververtegenwoordiging in de steekproef van de desbetreffende categorie van opleiding. De in de tabel vermelde p-waarden hebben betrekking op een chi-kwadraattest waarbij de populatieverdeling gebruikt werd om de verwachte frequenties te bepalen. Telkens gaat het om p-waarden die kleiner zijn dan een honderste. Dit betekent dat bij de vijf surveys voor het kenmerk 'behaald diploma' er telkens een significant verschil is tussen de steekproefverdeling en de populatieverdeling.

Tabel 2: Verschil (uitgedrukt in percentpunten) tussen de steekproefverdeling van het kenmerk 'behaald diploma' en de overeenstemmende populatieverdeling.

	1996	1997	1998	1999	2000
Lager onderwijs of geen	-8.7	-12.4	-12.7	-11.4	-10.6
Lager secundair onderwijs	5.0	8.2	5.1	4.9	3.8
Hoger secundair onderwijs	-1.0	-0.7	4.4	7.9	2.9
Niet-universitair hoger ond.	2.2	2.6	1.2	-1.8	3.2
Universitair onderwijs	2.6	2.2	2	0.4	0.8
	P<0.01	P<0.01	P<0.01	P<0.01	P<0.01

De belangrijkste vaststelling bij tabel 2 is ongetwijfeld de zware ondervertegenwoordiging van de laagste categorie van opleidingsniveau (lager onderwijs of geen). De oververtegenwoordiging van de categorie 'lager secundair onderwijs' is wellicht het gevolg van sociaal wenselijke antwoorden van lager opgeleiden die hun opleidingsniveau 'overschatten'. Deze oververtegenwoordiging compenseert die ondervertegenwoordiging van 'lager onderwijs of geen' zeker niet. De categorieën 'niet-universitair hoger onderwijs' en 'universitair onderwijs' worden op een uitzondering na telkens overschat. De vergelijking tussen de steekproefverdeling en populatieverdeling maakt duidelijk dat in de vijf surveys het opleidingsniveau systematisch wordt overschat. Er dient daarbij opgemerkt te worden dat dit wellicht samenhangt met (non)respons die gerelateerd is aan andere basisvariabelen zoals leeftijd en geslacht (cfr. infra).

### 2.2.2 Vergelijking tussen respondenten en niet respondenten

De vergelijking tussen de respondenten en zij die niet meewerkten aan het interview kan ook nog gebeuren op basis van informatie uit het steekproefkader en op basis van informatie die bekomen wordt door middel van het contactformulier.

#### - Op basis van informatie uit het steekproefkader

Het steekproefkader bevat voor alle eenheden van de populatie soms informatie over een beperkt aantal kenmerken. Indien deze informatie bij de steekproeftrekking mee geregistreerd wordt, dan kan deze informatie gebruikt worden om na te gaan of het (non)responsprofiel gerelateerd is aan deze kenmerken.

Voor alle getrokken steekprofeenheden van het survey van 2000 hebben we uit het steekproefkader informatie kunnen overnemen over geboortejaar en geslacht. Dit laat ons toe om de relatie na te gaan tussen beide achtergrondkenmerken en het (non)responsprofiel.

Tabel 3: (non)responsprofiel naar geslacht en leeftijd.

Geslacht	Man					Vrouw				
	Leeftijd	16-34	35-54	55-74	75-85	Tot-man	16-34	35-54	55-74	75-85
Afgew. Interviews	68.8	68.6	63.6	69.1	67.3	69.7	65.1	53.7	45.9	61.7
Weigeringen	14.6	15.3	21.9	23.8	17.2	17.1	22.5	36.1	40.2	26.2
Niet te contacteren	11.8	13.1	12.3	2.4	12.1	10.5	9.9	7.8	9.2	9.4
Andere nonrespons	4.8	3.0	2.2	4.7	3.5	2.6	2.5	2.4	4.6	2.7
Totaal (n)	328	398	269	42	1037	304	364	294	87	1049

In bovenstaande tabel worden de (non)responsprofielen weergegeven in de combinaties van de categorieën van geslacht en leeftijd. Uit een log-lineaire analyse van deze tabel blijkt dat er een significante relatie is tussen het (non)responsprofiel en geslacht (chi-kwadraat= 15.74, df=3,  $p < 0.01$ ). Onder controle van leeftijd zijn vrouwen iets gemakkelijker te contacteren maar weigeren ze meer dan mannen. Dit resulteert in een verschil tussen mannen en vrouwen van 5.6 procentpunten wat betreft het percentage afgewerkte interviews. Dit resultaat ligt niet in de lijn met de algemene tendens die in ander onderzoek wordt vastgesteld en waarbij mannen in mindere mate participeren aan interviews (Groves & Couper, 1998, 136). Een afwijking van deze algemene trend werd echter ook vastgesteld bij (non)responsanalyses van de Vlaamse data van het Belgisch verkiezingsonderzoek. Ook in dit onderzoek weigerden vrouwen meer dan mannen. Een mogelijke verklaring daarvoor werd gezocht in de geringere politieke kennis van vrouwen waardoor ze minder geneigd zijn om deel te nemen aan een interview over politieke onderwerpen. Verder blijkt uit de log lineaire analyse van de tabel ook dat er een significante partiële samenhang bestaat tussen leeftijd en (non)responsprofiel (chi-kwadraat= 40.62, df=9,  $p < 0.01$ ). Bij toenemende leeftijd stijgt het percentage weigeringen. Ouderen zijn echter wel gemakkelijker te contacteren. Een dergelijk leeftijdseffect is consistent met wat in ander onderzoek werd vastgesteld (Groves & Couper, 1998, 134). Opmerkelijk zijn de lage percentages afgewerkte interviews in de twee hoogste leeftijdscategorieën bij de vrouwen. Dit resulteert echter niet in een significant interactie-effect tussen leeftijd, geslacht en (non)responsprofiel (chi-kwadraat= 9.06, df=9,  $p = 0.43$ ). Dit betekent dat de relatie tussen (non)responsprofiel en leeftijd niet significant verschilt voor de mannen en de vrouwen.

Op basis van dezelfde informatie kan de vergelijking tussen respondenten en niet respondenten nog op een andere wijze worden benaderd. Zo kunnen we de categorieën 'Weigeringen', 'Niet te contacteren' en 'Andere nonrespons' samenvoegen in één categorie 'non-respons' en vergelijken met de respons (=afgewerkte interviews). Uiteindelijk willen we weten of zij die meewerkten (respons) verschillen in leeftijd en geslacht van diegene die niet meewerkten (non-respons).



Tabel 4: percentage respons en non-respons naar leeftijd en geslacht

		Respons	Non-respons
Man	16-34	16.7	13.9
	35-54	20.3	16.9
	55-74	12.7	13.2
	75-85	2.2	1.8
Vrouw	16-34	15.8	12.4
	35-54	17.6	17.1
	55-74	11.8	18.3
	75-85	2.9	6.3
Totaal		1345	741
Chi-kwadraat= 37.38, df=7, p<0.01			

Uit tabel 4 blijkt dat er een significant verschil is tussen de respons en de non-respons wat betreft de gecombineerde verdeling van geslacht en leeftijd. De verschillen tussen beide groepen zijn het grootst bij de jongste twee leeftijdscategorieën bij de mannen en de oudste twee leeftijdscategorieën bij de vrouwen. De verschillen bevestigen de eerder gemaakte vaststelling dat het vooral oudere vrouwen zijn die minder participeren aan het interview. Dit is ook in overeenstemming met de eerder gemaakte vaststelling dat lager opgeleiden minder meewerken aan een interview. Oudere vrouwen zijn immers lager opgeleid. De resultaten bevestigen ook de bevindingen van de (non)responsanalyses van de Vlaamse data van het Belgisch verkiezingsonderzoek.

De vastgestelde verschillen tussen respons en non-respons bevestigen dus in grote mate andere onderzoeksresultaten en zijn vrij aannemelijk. Toch zijn ze niet evident en dit omwille van de volgende reden. Bij de vervangingsprocedure werd er immers in de mate van het mogelijke voor gezorgd dat de vervanging van een effectieve steekproefeenheid gebeurt door een reserve steekproefeenheid van ongeveer dezelfde leeftijd. Dit betekent dat wanneer iemand weigert of niet kan gecontacteerd worden of om een andere reden niet kan deelnemen aan het interview hij of zij vervangen wordt door iemand met ongeveer dezelfde leeftijd. Blijkbaar resulteert deze vervangingsprocedure niet in een responsgroep die wat betreft leeftijd gelijk is aan de (non)responsgroep.

#### - Op basis van informatie van het contactformulier

Door middel van een contactformulier dat de interviewer dient in te vullen bij elk contact met een steekproefeenheid die hem werd toegewezen, kan eveneens bijkomende informatie bekomen worden over zowel de respondenten als diegene die niet meewerkten. Op het contactformulier noteert de interviewer ondermeer wanneer hij met een steekproefeenheid contact had en wat het resultaat ervan was. Aan de interviewer wordt ook gevraagd om op het contactformulier gegevens over het type woning, de staat van de woning en de woonomgeving van het steekproefelement te noteren. De registratie van dit soort gegevens steunt dus op de observatie en evaluatie van de interviewers. Dergelijke informatie is beschikbaar voor

alle gebruikte steekproefeenheden en kan dus gebruikt worden om respondenten te vergelijken met zij die niet meewerkten. In tabel 5 wordt gebruik gemaakt van de informatie over het type van de woning dat door middel van het contactformulier werd verzameld. Uit deze tabel blijkt dat er significante verschillen zijn wat betreft woningtype tussen respondenten, weigeraars, de categorie van de niet te contacteren steekproefeenheden en de restcategorie. Bij de respondenten woont 41% in een alleenstaande woning. Dit is beduidend meer dan in de andere categorieën. Slechts 11% van de respondenten woont in een meergezinswoning. Bij weigeraars en de groep niet te contacteren loopt dit op tot respectievelijk 19% en 28%. Dit laatste bevestigt de idee dat appartementsbewoners moeilijk te contacteren zijn. Uit dit alles blijkt dat het soort woning van de respondenten duidelijk verschillen vertoont met het type woning van zij die niet meewerkten. Wellicht gaat het niet alleen om verschillen in woningtype maar om meer algemene verschillen in leefsituatie die met die woningtypes verbonden zijn.

Tabel 5: Type woning per categorie van het responsprofiel

	Respondenten	weigeraars	Niet te contacteren	Andere non-respons
Alleenstaande woning	41.0	28.8	25.8	27.8
Half open bebouwing	18.7	14.9	14.0	11.1
Rijhuis	23.3	32.6	29.4	38.9
Meergezinswoning	11.7	19.1	28.1	14.8
Andere	6.0	4.6	2.7	7.4
totaal	1339	430	221	54

### 2.2.3 Vergelijking tussen respondenten en weigeraars

De vergelijking tussen respondenten en zij die niet meewerkten aan het interview op basis van informatie die kan worden overgenomen uit het steekproefkader en het contactformulier is doorgaans beperkt tot een paar klassieke achtergrondkenmerken. Om van de non-responsgroep meer inhoudelijke relevante informatie te bekomen kan men aan de weigeraars vragen om vooralsnog mee te werken aan een zeer kort interview. Dit is minder absurd dan op het eerste zicht lijkt. Binnen bepaalde surveys wordt 'refusal conversion' als een standaard onderdeel van de dataverzamelingsprocedure toegepast. Dit betekent dat men steekproefeenheden die weigeren om deel te nemen door een andere interviewer opnieuw gaat benaderen. Door deze werkwijze slaagt men erin om van een aantal oorspronkelijke weigeraars toch nog een interview af te nemen. De mogelijkheid van succesvol 'refusal conversion' maakt duidelijk dat weigeren geen permanente toestand. Men weigert in bepaalde omstandigheden om deel te nemen aan een bepaald survey, maar in ander omstandigheden (o.a. een ander geschikt tijdstip, andere interviewer,...) reageert men positief.

In het survey naar culturele verschuivingen in Vlaanderen 2000 werden bijna 35% (157/453) van alle weigeringen opnieuw telefonisch gecontacteerd. Het voornaamste objectief van deze telefonische interviews was om te controleren of de weigeraar effectief persoonlijk con-

tact had gehad met de interviewer en of de interviewer een man of een vrouw was. Daarnaast zijn in de mate van het mogelijk ook enkele vragen gesteld over een paar sociaal demografische kenmerken en werd het telefonisch interview afgesloten met het voorleggen van enkele uitspraken van een utilitair individualisme schaal. Het meten van utilitair individualisme bij de weigeraars kan verwondering wekken. In wat volgt wordt daar verder op ingegaan. Uiteindelijk werden 90 weigeraars bereid gevonden om mee te werken aan een kort telefonisch interview en voor deze groep hebben we voldoende informatie om de vergelijking te maken met de respondenten. De groep van weigeraars die meewerkt aan het telefonisch interview verschilt niet significant wat betreft leeftijd en geslacht van de groep weigeraars die ook het telefonisch interview weigert. Dit betekent echter niet noodzakelijk dat de telefonisch geïnterviewde 'weigeraars' een representatieve groep is van alle weigeraars. De selectie van de weigeraars die telefonisch werden benaderd, gebeurde immers niet op toevalsbasis. In onderstaande tabel wordt de vergelijking weergegeven tussen de respondenten en de weigeraars die meewerkten aan het korte telefonisch interview.

Tabel 6: vergelijking tussen respondenten en telefonisch geïnterviewde 'weigeraars'

Kenmerk	Respondenten N=1345	Weigeraars N=90
Leeftijd ( $\chi^2 = 381.13$ , $df=2$ , $p<0.01$ )		
<35 jaar	32.5	14.4
35-55 jaar	51.4	57.8
55 plus	16.1	27.8
Geslacht ( $\chi^2 = 79.31$ $df=1$ , $p<0.01$ )		
Man	51.9	40.0
Vrouw	48.1	60.0
Hoogst behaalde diploma ( $\chi^2 = 431.28$ , $df=2$ , $p<0.01$ )		
Geen of lager onderwijs	16.2	39.8
Lager secundair onderwijs	26.7	14.8
Hoger secundair onderwijs	33.9	33.0
Niet-universitair en universitair	23.2	12.5
Betaald werk ( $\chi^2 = 88.13$ $df=1$ , $p<0.01$ )		
Ja	54.6	42.0
Neen	45.4	58.0
Utilitair individualisme ( $\chi^2 = 148.71$ , $df=2$ , $p<0.01$ )		
Zwak	15.7	14.8
Matig	60.2	45.5
Sterk	24.1	39.8

De resultaten in tabel 6 bevestigen de eerdere evaluaties. Bij de weigeraars zijn er significant meer ouderen (55 plus), vrouwen, lager geschoolden en personen zonder betaald werk. Bemerk dat personen zonder een betaalde baan gepensioneerden, thuiswerkenden, zieken, etc.. kunnen zijn.

Wat het utilitair individualisme betreft stellen we vast dat weigeraars zich in meerdere mate situeren in de categorie ‘sterk utilitair individualisme’. Utilitair individualisme kan algemeen omschreven worden als de mate waarin men gericht is op de verwerving van materiële middelen en status en kan beschouwd worden als een uitdrukking op het individueel niveau van een algemeen maatschappelijk klimaat dat gekenmerkt wordt door eigenbelang en een geringe maatschappelijke verantwoordelijkheid. Men kan verwachten dat personen die utilitair individualistisch zijn ingesteld slechts in geringe mate bereid zullen zijn om mee te werken aan een interview. De deelname aan een interview is immers niet direct functioneel voor hun eigenbelang. Het significante verschil tussen respondenten en weigeraars dat in bovenstaande tabel wordt vastgesteld, bevestigt deze verwachting en is ook in overeenstemming met de resultaten van ander onderzoek waar werd vastgesteld dat panel respondenten die hoog scoren op utilitair individualisme schaal meer weigeren om mee te werken aan de volgende golf van een panel (Loosveldt & Carton, 2001). Dit alles wijst op het belang van utilitair individualistische overwegingen bij de beslissing om al dan niet deel te nemen aan een interview.

### 2.3 Conclusie

Bij de outputevaluatie van een gerealiseerde steekproef werden de twee bepalende factoren van de non-responsfout met name het non-respons percentage en het verschil tussen respondenten en zij die niet meewerkten, grondig geëvalueerd. De evaluaties van beide factoren bij de survey naar culturele verschuivingen in Vlaanderen 2000 maken duidelijk dat er wel degelijk sprake is van non-responsfout in dit survey. Het percentage steekprofeenheden dat niet meewerkt aan dit survey is niet verwaarloosbaar en bij een aantal inhoudelijk relevante kenmerken zijn er significante verschillen vastgesteld tussen respondenten en zij die niet meewerkten. De onder- en oververtegenwoordiging van bepaalde groepen zijn een bevestiging van wat ook in ander onderzoek werd vastgesteld. De conclusie is dan ook dat omwille van de non-responsfout de steekproef niet zomaar kan beschouwd worden als een representatieve steekproef. Een klassieke remedie voor dit probleem bestaat erin om de steekproef te wegen op basis van beschikbare populatie informatie. Eerder werd reeds gesteld dat deze informatie beperkt is. Het wegen van de steekproef lost zeker niet alle problemen op en het effect ervan mag zeker niet overschat worden. Niet alles wordt immers perfect gedetermineerd door de kenmerken op basis waarvan de weging gebeurt. Ook in deze context is voorkomen beter dan genezen. In de toekomst dient er in ieder geval meer aandacht besteed te worden aan de wijze waarop men de medewerking kan verkrijgen van laag geschoolden en oudere vrouwen.

### 3. Procesevaluatie van de gerealiseerde steekproef

De procesevaluatie van de gerealiseerde steekproef richt zich op de contact- en vervangingsprocedure en de wijze waarop deze procedure wordt toegepast. De algemene idee daarbij is dat interviewers maximale inspanningen moeten leveren om in contact te komen met de respondent en om deze te overtuigen om mee te werken. In ieder geval mag de procedure niet toelaten dat een steekproefeenheid te snel wordt vervangen. Indien men dit laatste wel doet dan komt men immers bij gemakkelijk te contacteren en te overtuigen respondenten terecht. Er wordt aangenomen dat deze groep verschilt van moeilijk te contacteren en weigerachtige respondenten.

Alvorens verder in te gaan op de contact- en vervangingsprocedure dient nog vermeld te worden dat met het oog op de kwaliteit van de verzamelde gegevens het beter is om met veel interviewers te werken die elk een beperkt aantal interviews afnemen in plaats van met weinig interviewers die elk veel enquêtes realiseren (Groves, 1989, Pickery, 2000). Omwille van die reden werd in het survey naar culturele verschuivingen in Vlaanderen 2000 contractueel vastgelegd dat interviewers maximaal 2 sets van telkens 10 interviews mochten afnemen. Uiteindelijk stonden 85 interviewers in voor de realisatie van 1345 interviews. De afspraak in verband met een maximum van 20 interviews per interviewer kon niet worden gehandhaafd. De afwijking kwam er nadat werd vastgesteld dat de vooropgezette timing niet kon worden gehaald. Een kwart van de interviewers interviewden meer dan 20 respondenten en hebben uiteindelijk 45.5% van alle interviews gedaan. Dit betekent dus dat een omvangrijk aandeel van alle interviews gerealiseerd wordt door een eerder beperkte groep interviewers.

Eerder werd reeds vermeld dat 35% van de weigeraars opnieuw gecontacteerd werd. Twee derde ervan stemde ermee in om deel te nemen aan een kort telefonisch interview. Deze vaststelling bevestigt de idee dat de groep van de harde weigeraars, namelijk diegene die steeds weigeren om deel te nemen aan een survey, eerder beperkt is. Ook ander follow-up onderzoek bij weigeraars komt tot de vaststelling dat mits de nodige tijd en financiële middelen er hoge(re) responscijfers kunnen worden gerealiseerd (Stoop en Louwen, 2000). Een en ander wijst erop dat de wijze waarop contacten worden georganiseerd en de inspanningen die geleverd worden om de effectieve steekproefeenheden te interviewen sterk bepalend zijn voor het (non)respons percentage dat men realiseert. Non respons is met andere woorden niet alleen een kwestie van respondentengedrag maar ook een kwestie van interviewergedrag en de wijze waarop de onderzoeker dit interviewergedrag ondersteunt. Op basis van een multilevel analyse werd trouwens vastgesteld dat er significante verschillen zijn ( $p < 0.001$ ) tussen de interviewers wat betreft de proportie afgewerkte interviews en dit onder controle van geslacht en leeftijd van de respondent. Zo slagen de twee beste interviewers<sup>1</sup> erin om samen 60 interviews af te nemen van 65 steekproefeenheden (95% respons). De twee slechtste interviewers daarentegen realiseren slecht 25 interviews op een totaal van 60 steek-proefeenheden een respons (40% respons).

---

1. De twee beste en twee slechtste interviewers werden geselecteerd op basis van de resultaten van de multilevelanalyse. Het gaat om de twee grootste en de twee kleinste residuele waarden op het interviewerniveau.

### 3.1 *Introductiebrief*

Algemeen wordt aangenomen dat het versturen van een introductiebrief naar de steekproefeenheden alvorens ze persoonlijk te contacteren, het eerste contact vlotter laat verlopen en de kans tot medewerking verhoogt. Experimenteel onderzoek heeft aangetoond dat de vermelding van wie het onderzoek organiseert en de inhoud van de brief belangrijke effecten hebben op het verkrijgen van medewerking (Groves & Couper, 1998). Wanneer men over een adressenlijst van de steekproefeenheden beschikt en het onderzoek georganiseerd wordt door een officiële instantie dan verdient het zeker aanbeveling om met een introductiebrief te werken. Daarbij dient men er wel voor te zorgen dat de periode tussen de ontvangst van de brief en het moment waarop de steekproefeenheid voor het eerst gecontacteerd wordt, niet te lang wordt. Onderzoek heeft aangetoond dat wanneer die periode te lang wordt, dit de medewerking doet dalen.

In het survey naar culturele verschuivingen in Vlaanderen 2000 werd met een introductiebrief gewerkt. De brief voor de effectieve steekproefeenheden werd centraal verstuurd vanuit de administratie Planning en Statistiek van het Ministerie van de Vlaamse Gemeenschap. De interviewers moesten zelf de introductiebrieven versturen naar de vervangingen. Er zijn aanwijzingen dat dit niet altijd gebeurde. Er werd vastgesteld dat vervangingen voor een zelfde respondent op dezelfde dag plaatsvonden. Bij deze vervangingen was het niet mogelijk om brieven te versturen. Verder zijn er weinig gegevens beschikbaar om het gebruik en het effect van de introductiebrief te evalueren. Maar globaal genomen mag worden aangenomen dat de brief een positief effect had zoals dat in ander onderzoek werd vastgesteld.

### 3.2 *Contactprocedure*

Uit onderzoek dat gericht is op het optimaliseren van de kans op contact met een steekproefeenheid blijkt dat gezinnen met kinderen, gezinnen met meer volwassenen en bejaarden gemakkelijker kunnen worden gecontacteerd. Contacten tijdens de avond zijn productiever dan contacten over dag. Dit is ook zo voor contacten tijdens het weekend. Dit impliceert onder andere dat bijkomende contacten 's avonds en tijdens de weekends er voor zorgen dat steekproefeenheden uit kleinere gezinnen worden gecontacteerd. (Groves, 1998; Purdon, Campanelli, & Sturgis, 1997). Wat de wijze van contacteren betreft (face-to-face of telefonisch) heeft eerder onderzoek uitgewezen dat interviewers die eerste contacten overwegend telefonisch doen niet per definitie minder goede resultaten bekomen (Carton, 1999). De resultaten van deze onderzoeken naar het effect en de efficiëntie van contactprocedures laten zich gemakkelijk vertalen in een concrete contactprocedure.

Voor het survey naar culturele verschuivingen in Vlaanderen 2000 kon het eerste contact zowel telefonisch als face-to-face gebeuren. Bij een telefonisch weigering was de interviewer verplicht om de respondent toch nog face-to-face te benaderen en hem alsnog te overtuigen om deel te nemen aan het onderzoek. De contactprocedure stipuleerde verder dat bij afwezigheid van de respondent er twee maal terug contact diende te worden opgenomen en dit op verschillende dagen en uren. Minstens één contactpoging moest in het weekend worden ondernomen en één op werkdagen na 18 uur.

Uit de informatie die verzameld werd door middel van het contactformulier blijkt dat 11.5% van de eerste contacten met de respondent telefonisch gebeurden. In meer dan de helft van de gevallen (57,7%) resulteert dit in een afspraak. In 13.7% van de telefonische eerste con-

tacten weigerde de respondent zijn medewerking. Dit is drie procentpunten meer dan in het geval van een persoonlijk eerste contact. Dit is geen bijzonder groot verschil maar in de context van (non)respons toch ook niet verwaarloosbaar. Bij 80% van deze telefonisch weigeringen is er informatie over een tweede persoonlijk contact. Slechts in een beperkt aantal gevallen leverde dit ook een interview of een afspraak op. De overgrote meerderheid (85%) bevestigt de eerdere telefonische weigering. Omwille van het kleine aantal telefonische contacten dat een weigering opleverde is de impact ervan op het uiteindelijk (non)response percentage gering. Het verdient echter geen aanbeveling om interviewers te stimuleren om eerst telefonisch contact op te nemen met de respondent.

Bijna 15% van de eerste contacten (telefonisch en face-to-face samen) resulteert in een afgewerkt interview (zie tabel 7). In meer dan een kwart van de gevallen wordt een afspraak gemaakt en in iets meer dan 10% bekommt men een weigering. Opmerkelijk is ook dat in iets meer dan 40% van de contacten men de respondent niet thuis aantreft. Het eerste contact is met andere woorden niet echt efficiënt om de respondent thuis te vinden. Bij nader toezien blijken 61% van die eerste contacten voor zes uur 's avonds plaats te vinden. Wanneer het eerste contact na zes uur plaats vindt, daalt het percentage 'niet te contacteren' met 13 procentpunten en verandert het percentage weigeringen niet. Wel worden er 10 procentpunten meer afspraken gemaakt. Een eerste contact na zes 's avonds heeft dus een hoger rendement.

Tabel 7: resultaat van de eerste drie contacten

	Eerste Contact	Tweede contact		Derde contact	
		Met	Zonder	Met	Zonder
Afgewerkt interview	14.8	78.5	14.3	78.7	10.7
Afspraak	26.4	9.4	22.4	8.6	22.0
Weigering	11.6	4.0	11.3	3.7	14.3
Niet te contacteren	41.3	6.7	47.4	6.4	48.6
Andere nonrespons	5.9	1.4	4.6	2.6	4.3
Totaal	2111	554	969	267	531

In tabel 7 is er ook informatie opgenomen met betrekking tot het tweede en het derde contact en dit opgesplitst naar het feit of het contact al of niet gebeurde op basis van een afspraak die gemaakt werd tijdens een vorig contact. Opmerkelijk is het gelijkaardig responsprofiel bij het tweede en het derde contact. Afspraken resulteren in bijna 80% van de gevallen in een interview, in bijna 10% moet een nieuwe afspraak worden gemaakt en iets meer dan 5% blijkt men ook na afspraak niet thuis te zijn. Iets minder dan 5% weigert nadat een afspraak werd gemaakt. Wanneer er geen afspraken worden gemaakt dan is het responsprofiel vergelijkbaar met het eerste contact. Ook hier blijkt terug dat de steekproefeenheden moeilijk te contacteren zijn. De (non)responsprofielen die in tabel 7 worden weergegeven zijn erg vergelijkbaar met wat in ander onderzoek werd vastgesteld (Carton, 1999) Na het eerste contact is 23% van het totaal aantal gerealiseerde interviews afgewerkt. Na het eerste en het tweede contact samen is dit tweede derde van de interviews en na het derde contact zijn 86% van de interviews gerealiseerd.

Aan de interviewers werd gevraagd dat wanneer het niet mogelijk was om de steekprofeenheid te contacteren, minstens eenmaal opnieuw contact op te nemen na 18.00 uur. Bij bijna 60% van de steekprofeenheden die bij het eerste contact voor 18.00 u niet gecontacteerd konden worden, gebeurde het tweede contact opnieuw voor 18.00u. Meer nog, in de helft van de gevallen waarbij zowel het eerste als het tweede contact plaats vond voor 18.00u en waarbij de steekprofeenheden tijdens beide contacten niet gecontacteerd konden worden, gebeurt het derde contact eveneens voor 18.00uur. In deze laatste situatie wordt de contactprocedure niet gevolgd. Bij 13% van de steekprofeenheden die uiteindelijk in de categorie 'niet te contacteren' terechtkomen, was er geen enkel contact na zes uur. Er zijn ook significante verschillen tussen interviewers wat betreft het aantal contacten na zes uur 's avonds. Dit betekent dus dat sommige interviewers significant meer tijdens de avonden interviewen. Dit kan als gevolg hebben dat ze ook andere groepen van respondenten interviewen. Volgens de contactprocedure moest er minstens een contact zijn tijdens het weekend wanneer de steekprofeenheid niet kon gecontacteerd worden. Interviewen tijdens de weekends lijkt niet zo populair te zijn. Slechts bij iets meer dan een derde van de steekprofeenheden (37%) is er een contact geweest tijdens het weekend. Ook hier zijn er significante verschillen tussen de interviewers. Opmerkelijk is ook dat bij bijna een vierde van de steekprofeenheden (23.6%) die definitief als niet te contacteren worden beschouwd er geen contact was tijdens het weekend.

### 3.3 *Vervangingsprocedure*

In het kader van het survey naar culturele verschuivingen in Vlaanderen 2000 werd in verband met de vervangingen het volgende contractueel bepaald: 'Bij voorkeur nemen de interviewers telkens contact op met de opdrachtnemer voor het bekomen van een vervangadres, m.a.w. vervangadressen worden niet onmiddellijk bij de start van het veldwerk meegegeven aan de interviewers.' Uiteindelijk werd met het marktonderzoeksbureau afgesproken dat de interviewers bij aanvang van hun interviewerwerk geen vervangadressen kregen. Vervangadressen konden pas worden doorgegeven nadat de effectieve adressenlijst volledig was afgewerkt. Van bij het begin van het veldwerk werd deze procedure zowel door het bureau als door de interviewers als zeer 'vertragend' beschouwd.

Ook hier zijn er aanwijzingen dat de vervangingsprocedure niet altijd werd gevolgd. Op basis van informatie die verzameld werd via het contactformulier kan worden nagegaan wanneer de interviewers starten met het gebruik van de vervangadressen. Indien de vervangingsprocedure correct wordt toegepast, dient het tijdstip waarop contact werd opgenomen met de eerste vervanging later te vallen dan de datum van het laatste contact met de laatste afgewerkt effectieve adres. Bij drie op tien van de interviewers vinden we aanwijzingen dat zij reeds over vervangadressen beschikten nog vooraleer het definitief resultaat gekend was van elke effectieve steekprofeenheid.

Alles lijkt erop te wijzen dat wanneer interviewers reeds bij aanvang van hun intervieweropdracht over vervangingsadressen zouden beschikken, ze gemakkelijker en sneller van die vervangingsadressen zouden gebruik maken. Dit is niet in overeenstemming met het uitgangspunt dat er maximale inspanningen moeten worden geleverd om de effectieve steekprofeenheden te interviewen. Het maximaliseren van deze inspanningen lijkt een noodzakelijke voorwaarde om zowel een hoog responspercentage te realiseren als om het verschil tussen zij die meewerken en zij die dat niet doen te minimaliseren.



### 3.4 Conclusie

Bij de procesevaluatie van de gerealiseerde steekproef werd vastgesteld dat interviewers heel wat inspanningen moeten leveren om steekproefeenheden te contacteren en te overtuigen om aan het interview mee te werken. Sommige interviewers hebben daarbij meer succes dan anderen. Er werd ook vastgesteld dat de dataverzameling niet altijd volgens de vooropgestelde procedure is verlopen. Een en ander wijst erop dat de contactprocedure kan geoptimaliseerd worden. De efficiëntie van het eerste contact kan wellicht ook worden verbeterd door een betere timing ervan. Meer eerste contacten tijdens de avonduren en het beter spreiden van de contacten overdag en 's avonds zijn daarbij belangrijk. Het werd ook duidelijk dat het maximum aantal contacten niet beperkt kan worden tot drie. Bijna 15% van de interviews gebeurt trouwens niet tijdens de eerste drie contacten.

Interviewers moeten blijvend gemotiveerd worden om maximale inspanningen te leveren om effectieve steekproefeenheden te interviewen. Dit kan onder andere gebeuren tijdens een interviewtraining. Interviewers kunnen beter worden voorbereid en getraind hoe ze respondenten moeten contacteren en hoe ze respondenten moeten overtuigen om mee te werken. Dit belangrijk onderdeel van zijn taak is immers duidelijk verschillend van het interviewen in enge zin. De taak van de interviewer tijdens het interview is doorgaans veel meer ondersteund vanuit het onderzoeksopzet (o.a. vragenlijst met instructies) dan zijn opdracht tijdens de introductie aan de deur. Het contacteren en overtuigen van de respondent is minder gestructureerd dan wat er moet gebeuren tijdens het interview. Er wordt zelf meer en meer beklemtoond dat de interviewer bij het eerste contact met de respondent zich moet aanpassen aan de wijze waarop de respondent reageert. Deze idee van 'tailoring' tijdens het eerste contact ligt niet in de lijn van de sterk gestructureerde aanpak tijdens een gestandaardiseerd interview. Deze flexibele opstelling van de interviewers vereist een uitgebreidere training.

### 4. De outputevaluatie van de geregistreerde antwoorden

Zoals eerder gesteld richt de outputevaluatie van de geregistreerde antwoorden zich op meetfouten. Meetfouten vinden hun oorsprong onder andere in inadequaat interviewgedrag en respondentengedrag. Inadequaat interviewgedrag betekent dat de interviewer het antwoordproces beïnvloedt waardoor er interviewereffecten optreden bij de geregistreerde antwoorden. Het optreden van interviewervariantie is een concrete indicatie van dergelijke interviewereffecten. Interviewervariantie betekent dat een proportie van de variabiliteit in de geregistreerde antwoorden op een vraag kan 'verklaard' worden door de interviewer. De aanwezigheid van interviewervariantie wijst er ook op dat een belangrijke veronderstelling die gemaakt wordt bij het verzamelen van data door middel van gestructureerde interviews, niet aannemelijk is. Het gaat om de veronderstelling dat interviewers die vergelijkbare groepen respondenten interviewen ook vergelijkbaar antwoorden registreren. In feite mag het er niet toe doen of een respondent door interviewer A dan wel door interviewer B wordt geïnterviewd. De evaluatie van interviewervariantie is in die zin een belangrijk onderdeel van de outputevaluatie. Significante interviewervariantie is een sterke indicatie van meetfouten. Meetfouten kunnen ook hun oorsprong vinden in het antwoordgedrag van de respondenten. Respondenten moeten voldoende cognitieve en communicatieve vaardigheden hebben en voldoende gemotiveerd zijn om hun rol van betrouwbaar respondent op te nemen. Dit betekent dat ze bereid zijn en in staat zijn om voldoende inspanningen te leveren om het cogni-

tief proces, dat bij het beantwoorden van de vragen doorlopen moet worden, optimaal te laten verlopen. Het optimaal verloop van het cognitief proces is een voorwaarde voor volledige, nauwkeurige en betrouwbare antwoorden. Of anders uitgedrukt, een gebrekkige motivatie en onvoldoende capaciteit moeten normaliter resulteren in meetfouten waarvoor de respondent verantwoordelijk is. Door middel van het rapport van de interviewer kan men informatie bekomen om op indirecte wijze na te gaan of aan deze voorwaarde voldaan is. In het rapport van de interviewer wordt aan de interviewer gevraagd om – nadat het interview volledig is afgewerkt – een aantal feitelijke gegevens over het interview te noteren (bv. i.v.m. de aanwezigheid van derden) en om de wijze waarop het interview verliep, te evalueren. Daarbij kan onder andere gevraagd worden naar een evaluatie van de motivatie en de bekwaamheid van de respondent.

In het survey naar de culturele verschuivingen in Vlaanderen 2000 werd aan de interviewer gevraagd om de motivatie van de respondent te beoordelen op een vijfpuntschaal gaande van ‘zeer gemotiveerd’ tot ‘zeer terughoudend’. Slechts 2% van de respondenten kreeg een beoordeling van ‘terughoudend’ of ‘zeer terughoudend’. Voor de beoordeling van de capaciteit van de respondent om de vragen te beoordelen werd een zes-puntschaal gebruikt: van ‘zeer hoog’ tot ‘totaal onvoldoende’. Bij slechts 1.8% van de respondenten was de capaciteit – volgens de interviewer – ‘zeer gering’ of ‘totaal onvoldoende’. Op basis van deze beoordelingen moeten we concluderen dat de overgrote meerderheid van de respondenten gemotiveerd en capabel was om de vragen te beantwoorden en dat ze dus voldeden aan de basisvoorwaarde voor een interview. De eerder vastgestelde selectieve non-respons van onder andere lager opgeleiden is daar wellicht niet vreemd aan. In vergelijkbaar onderzoek waar dezelfde beoordelingen gebeurden was de groep respondenten die niet voldoende gemotiveerd en bekwaam was beduidend groter (motivatie 4.9% en bekwaamheid 6.2%) (Loosveldt & Carton, 1997). In dit vergelijkbaar onderzoek werd ook vastgesteld dat de beoordelingen door de interviewer conform de verwachtingen gerelateerd zijn aan de datakwaliteit. Zo worden er significant meer ‘weet niet’ antwoorden geregistreerd bij de respondenten met geringe motivatie en beperkte bekwaamheid.

In wat volgt worden een aantal concrete uitingen van het antwoordgedrag op een directere wijze geëvalueerd. Er is daarbij ook aandacht voor de mogelijke impact van de interviewer en het optreden van interviewervariantie.

#### 4.1 *Item non-respons*

Item non-respons kan het gevolg zijn van het niet vlot en optimaal verlopen van het antwoordproces. De respondent kan een vraag beoordelen als ‘te moeilijk’, ‘te bedreigend’ of ‘niet interessant’. Het kan ook dat een vraag de respondent in verlegenheid brengt. Om dergelijke problemen op te lossen kan de respondent weigeren om een vraag te beantwoorden of antwoorden met ‘weet niet’ of ‘geen mening’. Telkens gaat het om niet substantieel antwoord. Dergelijke antwoorden worden in een analyse doorgaans behandeld als item non-respons. Dit betekent dat de elementen met een dergelijk antwoord niet in de analyse worden opgenomen. Het is duidelijk dat wanneer deze groep niet vergelijkbaar is met diegene die de vragen wel hebben beantwoord, dit een bijkomende vertekening tot gevolg kan hebben. Voor 2.5% van de respondenten is er geen antwoord genoteerd bij de vraag naar het hoogst behaalde diploma of de huidige studierichting. Wanneer daaraan de niet erg bruikbare categorie ‘Ander’ wordt toegevoegd dan hebben we voor 3.5% van de respondenten geen bruik-

bare informatie over opleidingsniveau. Dit is niet bijzonder veel maar voor dit belangrijk achtergrondkenmerk toch ook niet verwaarloosbaar. Opmerkelijk is ook dat indien de respondent een partner heeft en dit is in twee derde van de gevallen, we op één uitzondering na informatie hebben over het hoogst behaalde diploma van deze partner. De informatie over het hoogst behaalde diploma van de partner is bijgevolg vollediger dan de informatie over het opleidingsniveau van de respondent zelf.

Bij de meer bedreigende vraag naar het netto maandelijks gezinsinkomen ligt het percentage item non-respons hoger dan bij de vraag naar opleiding. Een derde van de respondenten beantwoordt de open inkomensvraag niet. Bij de respondenten die de open inkomensvraag niet beantwoorden moest een gesloten inkomensvraag gesteld worden. Uiteindelijk beantwoordt 11% van de respondenten ook de gesloten inkomensvraag niet. Voor analyses waarbij de variabele inkomen wordt gebruikt is dit zeker een niet onbelangrijke beperking. Wanneer opleiding en inkomen samen in een analyse worden opgenomen dan worden 13% van de respondenten omwille van item non-respons op één van beide kenmerken niet gebruikt.

Bij vragen die wat moeilijker zijn en waarbij de respondent een grotere inspanning moet leveren om ze te beantwoorden mag men verwachten dat er iets meer item non-respons optreedt dan bij een doorsnee vraag. De vragenlijst van het survey naar culturele verschuivingen in Vlaanderen bevat weinig of geen moeilijk vragen. Eén van de moeilijkere onderdelen van de vragenlijst bestond erin om uit een lijst van 24 maatschappelijke problemen (o.a. verkeersdruk, de vreemdelingen, de milieuvervuiling, ...) er vijf te selecteren die de respondent het belangrijkste vond. Vervolgens moesten er drie van de vijf geselecteerde problemen gerangschikt worden naar belangrijkheid. De item non-respons bij deze vragen is zeer klein. Deze taak wordt door 97.25% van de respondenten perfect uitgevoerd. De 'meest' voorkomende fout (bij 1.9% van de respondenten) bestaat erin dat men slechts drie of vier problemen selecteert in plaats van 5. Deze geselecteerde problemen worden wel nog netjes geordend naar belangrijkheid. Dit betekent dat we ook bij deze respondenten genoeg geen item non-respons hebben bij dit onderdeel van de vragenlijst. In feite kan er gesteld worden dat slechts 1% van de respondenten problemen heeft bij het uitvoeren van deze taak. In de vragenlijst was nog een tweede selectie en ordeningsvraag opgenomen. Aan de respondent werd gevraagd om uit een lijst van 13 de drie meest bedreigende vormen van milieuverontreiniging te selecteren en te ordenen. Ook hier is het aantal problemen miniem: 98.4% van de respondenten voert de taak perfect uit. Er is een significante samenhang tussen het al of niet perfect uitvoeren van de eerste (i.v.m. maatschappelijke problemen) en de tweede (i.v.m. vormen van milieuverontreiniging) selectie en ordeningstaak. Toch is het niet zo dat de problemen bij de uitvoering van beide taken sterk cumulatief zijn: slechts 11% van de respondenten die de eerste taak niet perfect uitvoert, voert ook de tweede taak niet perfect uit. Wel is het zo dat het al of niet perfect uitvoeren van één of van beide selectie en ordeningstaken – zoals verwacht – significant samenhang met de beoordeling door de interviewer van de capaciteit van de respondent om de vragen te beantwoorden (chi-kwadraat=15.9, df=3, p<0.01). Bijna één op vijf (18.9%) van de respondenten die één of beide taken niet perfect uitvoert, krijgt de beoordeling 'gering', 'zeer gering' of 'totaal onvoldoende'. Bij diegene die beide taken perfect uitvoeren is dit slechts 7%.

Een 'geen mening' antwoord bij vragen waar deze antwoordmogelijkheid niet expliciet wordt aangeboden kan ook wijzen op een gebrek aan interesse of motivatie bij het beantwoorden van vragen. Respondenten houden zich immers aan de algemene regel die in nor-

male conversatie wordt gehanteerd en die stelt dat wanneer er een vraag gesteld wordt, er een antwoord wordt gegeven. Een 'geen mening' antwoord wordt in een survey interview dan ook niet frequent gegeven wanneer het niet tot de antwoordmogelijkheden behoort. In de vragenlijst naar culturele verschuivingen in Vlaanderen 2000 worden een aantal batterijen van uitspraken gebruikt met een 5 punt antwoordschaal van 'helemaal eens' tot 'helemaal oneens'. De 'geen mening' antwoordmogelijkheid staat niet op de antwoordkaart maar is wel voorzien op de vragenlijst. Bij drie batterijen met uitspraken (politieke machteloosheid, eigen levenswijze/burgerzin en milieu) met in het totaal 25 uitspraken geven 90% van de respondenten geen enkel 'geen mening' antwoord, 5% doet dit een enkele keer en 2.5% twee keer.

Deze resultaten bevestigen de algemene trend van het gebruik van 'geen mening' antwoord en wijzen erop dat het beantwoorden van deze items met behulp van een antwoordkaart vrij vlot verloopt. Indien men de 'geen mening' antwoorden bij deze items nog wil verminderen dan kan dit wellicht door de interviewer de instructie te geven dat ze er bij de respondent sterker moeten op aandringen om een van de antwoordmogelijkheden te kiezen die op de antwoordkaart staat.

Een 'weet niet' antwoord bij een opinie vraag is net zoals een 'geen mening' antwoord een niet substantieel antwoord die in analyses doorgaans niet gebruikt wordt. Een 'weet niet' antwoord kan dus ook beschouwd worden als een vorm van item non-respons. Het gebruik ervan kan een indicatie zijn van moeilijkheden in het antwoordproces. In het survey naar culturele verschuivingen in Vlaanderen 2000 is er een vraag opgenomen naar het vertrouwen in organisaties en instellingen. Aan de respondenten werd gevraagd om van 18 instellingen en organisaties te zeggen hoeveel vertrouwen ze er in hebben. Ze maakten daarbij gebruik van een 5-punt schaal ('zeer veel' tot 'zeer weinig'). De antwoordmogelijkheid 'weet niet' werd niet in de vraagtekst vermeld en stond ook niet vermeld op de antwoordkaart. De 'weet niet' antwoordmogelijkheid is bij elke organisatie en instelling wel vermeld in de vragenlijst. Met uitzondering van de Europese Commissie mag men aannemen dat de respondenten de meeste instellingen en organisaties kennen (o.a. politie en rijkswacht, het onderwijs, het gerecht, de Vlaamse regering, ...). Toch is een 'weet niet' antwoord bij deze vragen plausibel. In vergelijking met het geven van 'geen mening' antwoorden (zie vorige paragraaf), worden er dan ook meer 'weet niet' antwoorden geregistreerd. Vier vijfde van de respondenten geeft geen enkel 'weet niet' antwoord, 10% doet dit één enkele keer en 10% doet dit 2 keer of meer. Het hoogste percentage weet niet antwoorden (9.6%) wordt gegeven bij 'de Waalse politieke partijen'. Op de tweede plaats komt de Europese Commissie met 6.3%. Opmerkelijk daarbij is dat de 'Waalse politieke partijen' ook de hoogste gemiddelde score op de vertrouwensschaal behalen en dus op het minst vertrouwen kunnen rekenen. Er is een significante ( $p < 0.01$ ) en sterke positieve correlatie ( $r = 0.50$ ) tussen het geven van 'weet niet' antwoorden en het geven van 'geen mening' antwoorden op de items die in de vorige paragraaf werden besproken. Verder is er ook een significante ( $p < 0.01$ ) en een redelijke sterk negatieve correlatie ( $r = -0.24$ ) tussen het geven van 'weet niet' antwoorden en het gebruik van de middencategorie van de vijf punt schaal. Hoe meer 'weet niet' antwoorden men geeft hoe minder men gebruik maakt van de middencategorie. Zowel bij het aantal 'weet niet' antwoorden als bij het gebruik van de middencategorie zijn er significante verschillen tussen de interviewers. Dit blijkt uit een multilevel analyse waarbij leeftijd geslacht en opleidingsniveau als controle variabelen werden opgenomen (weet niet antwoorden:  $z = 2.43$ ,  $p = 0.02$ ; middencategorie:  $z = 2.07$ ,  $p = 0.04$ ). Een en ander wijst er dus op

dat bij het beantwoorden van de vragen naar vertrouwen in instellingen en organisaties er zich wel wat problemen voordoen en dat de interviewers daar op een verschillende wijze mee omgaan waardoor de betekenis van een 'weet niet' antwoord niet zo eenduidig is. Item non-respons blijkt dus niet alleen een kwestie van antwoordgedrag te zijn maar wordt ook mede bepaald door interviewergedrag.

#### 4.2 'Weet niet' antwoorden bij kennis vragen

Het gebruik van 'weet niet' antwoorden hoeft niet altijd te wijzen op moeilijkheden in het antwoordproces. Bij kennis vragen is een 'weet niet' antwoord een zeer aannemelijk antwoord. In het survey naar culturele verschuivingen in Vlaanderen 2000 werd op summier wijze de kennis in verband met milieubeleid gemeten. Aan de respondent werd gevraagd om voor vier afkortingen te zeggen waarvoor ze staan (OVAM, CFK, MIRA, KGA). Bij meer dan een derde van de respondenten (36.9%) wordt op alle vier de vragen een 'weet niet' antwoord genoteerd. Het gemiddeld aantal weet niet antwoorden bedraagt 2.73. Er zijn echter significante verschillen tussen de interviewers wat betreft het aantal 'weet niet' antwoorden op deze kennisvragen (multilevel analyse met geslacht leeftijd en onderwijs als controlevariabele:  $z=3.60$ ,  $p<0.01$ ). De significante verschillen tussen de interviewers verklaren 7% van de variantie in het aantal 'weet niet' antwoorden op deze kennisvragen. Interviewers verschillen blijkbaar in hun beoordeling van de antwoorden van de respondenten en sommige omcirkelen blijkbaar sneller dan andere 'de weet' niet categorie die op de vragenlijst voorzien was. Interviewers zouden in ieder geval meer instructies moeten krijgen over hoe ze moeten omgaan met de antwoorden van de respondenten op deze vragen.

#### 4.3 Het gebruik van extreme antwoordmogelijkheden

Bij heel wat vragen in een survey interview wordt gebruik gemaakt van vijf punt schalen. Enkele van deze veel gebruikte schalen zijn: helemaal eens tot helemaal oneens, zeer tevreden tot zeer ontevreden, zeer veel vertrouwen tot helemaal geen vertrouwen en zeer belangrijk tot helemaal niet belangrijk. Dergelijke schalen zijn bedoeld om zowel informatie over de richting als over de sterkte van de opinie te bekomen. Het gebruik ervan veronderstelt wel dat de respondenten de nodige cognitieve inspanningen leveren om dit onderscheid te maken. Uit interactieanalyse van items waarbij dergelijke vijf punt schalen gebruikt worden, blijkt dat een van de antwoordstrategieën er inderdaad in bestaat om eerst de richting te bepalen om vervolgens uitspraak te doen over de sterkte (Loosveldt, 1985). Het is belangrijk dat de respondenten bij het beantwoorden van vragen waarbij een vijf punt schaal gebruikt wordt alle antwoordmogelijkheden in overweging nemen. Men mag verwachten dat zelfs respondenten die geen uitgesproken opinie hebben bij het beantwoorden van langere lijsten van uitspraken, toch af en toe een van de extreme antwoordmogelijkheden gebruiken. Indien dit niet het geval is dan kan men aannemen dat de respondent weinig inspanningen levert om genuanceerd te antwoorden.

Een kwart van de respondenten gebruikt geen enkele keer een extreme antwoordmogelijkheid (helemaal eens of helemaal oneens) bij de drie lijsten van uitspraken (samen 25 uitspraken) die bij de bespreking van het geven van geen mening antwoorden reeds aan bod kwamen. Bij de iets kortere lijst waarmee men peilt naar het vertrouwen in 18 instellingen en organisaties, geeft meer dan een derde van de respondenten (35.2%) nooit een extreem

antwoord (zeer veel of zeer weinig). Wie weinig extreme antwoorden geeft bij het peilen naar vertrouwen geeft ook weinig extreme antwoorden bij de drie lijsten met uitspraken die iets verder in de vragenlijst volgen. Tussen beide is er een significante ( $p < 0.01$ ) en sterk positieve correlatie ( $r = 0.46$ ). Een kwart van de respondenten geeft op beide momenten slechts één of geen enkel extreem antwoord. Dit lijkt erop te wijzen dat een aantal respondenten bij het beantwoorden van deze vragen systematisch een driepunt schaal hanteren in plaats van een vijfpunt schaal. Door deze feitelijke beperking in de categorisatie van de antwoordschaal worden correlaties tussen variabelen afgezwakt (De Wit, 1994). Het gebruik van een driepunt schaal impliceert in feite dat er informatie is over de richting maar niet over de sterkte van de opinie. Blijkbaar leveren sommige respondenten onvoldoende inspanningen om hun opinie te preciseren en worden ze door de interviewer onvoldoende gestimuleerd om dit te doen. Het is opmerkelijk dat er significante verschillen zijn tussen de interviewers wat betreft het registreren van extreme antwoorden bij de drie lijsten met uitspraken (multilevel analyse met geslacht leeftijd en onderwijs als controlevariabele:  $z = 2.72$ ,  $p < 0.01$ ) maar dat dit niet het geval is bij het peilen naar vertrouwen in instellingen en organisaties (multilevel analyse met geslacht, leeftijd en onderwijs als controlevariabele:  $z = 0.59$ ,  $p = 0.55$ ). Bij sommige vragen kunnen interviewers blijkbaar wel een invloed hebben op het al of niet frequent gebruik maken van extreme antwoordmogelijkheden.

#### 4.4 Interviewvariantie bij inhoudelijke variabelen

Bij de vraag naar het vertrouwen in 18 instellingen en organisaties werd reeds vastgesteld dat de interviewers verschillen wat betreft het gebruik van de 'weet niet' antwoordcategorie en de middencategorie van de antwoordschaal. Men kan zich de vraag stellen in welke mate er ook interviewereffecten optreden bij de inhoudelijke variabelen die men op basis van deze informatie kan construeren. Op basis van een factoranalyse kunnen 11 van de 18 variabelen gereduceerd worden tot drie factoren: het vertrouwen in politieke instellingen (Vlaamse regering, Vlaamse politieke partijen, Vlaams parlement, Belgisch parlement, Belgische regering en Waalse politieke partijen), het vertrouwen in de administratie (Vlaamse en gemeentelijke administratie) en het vertrouwen in de traditionele instellingen (kerk, koning en het leger). Bij geen enkele van deze drie factoren was de interviewervariantie significant (multilevel analyse met geslacht leeftijd en onderwijs als controlevariabele: vertrouwen in politieke instellingen:  $z = 0.70$ ,  $p = 0.48$ ; vertrouwen in de administratie:  $z = 1.58$ ,  $p = 0.11$ ; vertrouwen in traditionele instellingen:  $z = 0.31$ ,  $p = 0.76$ ). Er zijn dus geen significante verschillen tussen de interviewers wat betreft het vertrouwen in instellingen en organisaties dat ze geregistreerd hebben bij de respondenten.

Een van de batterijen van uitspraken die gebruikt werd bij de analyse van de 'geen mening' antwoorden had betrekking op politieke machteloosheid. Bij deze uitspraken werden weinig problemen gesignaleerd. Op basis van vijf<sup>2</sup> van de negen uitspraken kan een (redelijk) betrouwbare schaal van politieke machteloosheid geconstrueerd worden (Cronbach's

- 
2. De uitspraken zijn: (1) Gaan stemmen heeft geen zin, de partijen doen toch wat ze willen. (2) De politieke partijen zijn alleen maar geïnteresseerd in mijn stem en niet in mijn mening. (3) Als er mensen zoals ik aan de politici hun opvattingen laten weten, dan zullen zij daar rekening mee houden. (4) Er stemmen zoveel mensen bij de verkiezingen dat mijn stem er niet toe doet. (5) Mensen zoals ik hebben wel degelijk invloed op wat de overheid doet.

Alpha= 0.73). De interviewervariantie bij de op basis van deze schaal gemeten politieke machteloosheid is niet significant. Er zijn dus geen significante verschillen tussen de interviewers wat betreft de bij de respondenten geregistreerde politieke machteloosheid (multilevel analyse met geslacht, leeftijd en onderwijs als controlevariabele:  $z=0.78$ ,  $p=0.21$ ).

De twee zonet gepresenteerde analyses in verband met vertrouwen in instellingen en organisatie enerzijds en politieke machteloosheid anderzijds wekken de indruk dat interviewervariantie bij inhoudelijk variabelen niet echt een probleem is. Om dit beeld wat te nuanceren presenteren we twee voorbeelden waarbij wel interviewervariantie wordt vastgesteld. Het eerste voorbeeld betreft het meten van utilitair individualisme. Op basis van zes uitspraken<sup>3</sup> kan een (zeer) betrouwbare schaal worden geconstrueerd die utilitair individualisme meet (Cronbach's Alpha=0.86). Onder controle van geslacht, leeftijd en onderwijsniveau is de interviewervariantie significant ( $z=2.85$ ,  $p<0.01$ ). Dit betekent dat sommige interviewers significant hogere of lagere scores van utilitair individualisme bekomen dan andere. Dit kan erop wijzen dat sommige interviewers het antwoordgedrag van de respondenten beïnvloeden door de wijze waarop ze de uitspraken introduceren en voorlezen door de wijze waarop ze reageren op de antwoorden van de respondent.

Het tweede voorbeeld van significante interviewervariantie heeft betrekking op vragen naar feitelijk gedrag. Meer in het bijzonder gaat het om de vraag naar het aantal uren dat men gewoonlijk vrij heeft tijdens een week- of werkdag en een identieke vraag naar het aantal uren dat men gewoonlijk vrij heeft tijdens een zaterdag of een vrije dag. De referentieperiode bij deze vragen was telkens de voorbije maand en er werd ook vermeld wat niet als vrije tijd kon beschouwd worden. Het gaat hier om zeer moeilijk te beantwoorden vragen. De respondent moet aandacht hebben voor de referentieperiode, de definitie van vrije tijd en het 'juiste antwoord' weet hij niet en moet hij dus schatten. De kans is vrij groot dat bij deze vragen de respondent te kennen heeft dat hij daar niet onmiddellijk kan op antwoorden en dat de interviewer de vragen verder moet toelichten en dat hij de respondent moet stimuleren om toch een aantal te noemen. De kans dat de interviewer het antwoord systematisch in een bepaalde richting stuurt is vrij groot. Dit wordt bevestigd door de significante interviewervariantie die bij beide vragen aanwezig is (multilevel analyse met geslacht leeftijd en onderwijs als controlevariabele: uren vrije tijd werkdag:  $z=3.05$ ,  $p<0.01$ ; uren vrije tijd vrije dag:  $z=3.51$ ,  $p<0.11$ ). De verschillen tussen de interviewers verklaren bij 'vrije uren werkdag' 5.3 % van de variantie en bij de variabele 'vrije uren vrije dag' is dit 7.6%. Dit zijn indicaties dat de interviewers het antwoordproces bij deze vragen op een systematische wijze beïnvloeden.

- 
3. De uitspraken zijn: (1) De mensheid, onze naasten, solidariteit.. wat een onzin allemaal, iedereen moet eerst voor zichzelf zorgen en zijn belangen verdedigen. (2) Men moet steeds zijn eigenbelang nastreven, en zich niet te veel van anderen aantrekken. (3) Wat telt zijn geld en macht, al de rest zijn praatjes. (4) Het nastreven van persoonlijk succes is belangrijker dan te zorgen voor een goede verstandhouding met anderen. (5) In de samenleving kan men maar beter eerst voor zichzelf zorgen. (6) Het is belangrijk om in de eerste te streven naar een vooraanstaande positie voor zichzelf.

#### 4.5 Conclusie

De outputevaluatie van de geregistreerde antwoorden levert een gemengd beeld op. De meeste respondenten zijn voldoende gemotiveerd en in staat om de vragen die gesteld worden tijdens het interview te beantwoorden. De respondenten beantwoorden naar best vermogen de meeste vragen. Op een paar niet onbelangrijke uitzonderingen na (opleiding en inkomen), is de item non-respons ('Geen mening', 'Weet niet') gering. Toch is niet altijd duidelijk wat de juiste betekenis van deze niet substantiële antwoorden is. Er zijn ook aanwijzingen dat er bij het beantwoorden van sommige vragen bepaalde antwoordtendensen optreden (bv. Het niet gebruiken van extreme antwoordmogelijkheden). De outputevaluatie maakt ook duidelijk dat de interviewers behoorlijk ingevulde vragenlijsten afleveren. Het zijn professionele interviewers die weten hoe men een vragenlijst moet invullen en die weten dat van hen verwacht wordt dat ze correct ingevulde vragenlijsten afleveren. Er zijn echter aanwijzingen dat ze bij bepaalde vragen het antwoordgedrag van de respondent beïnvloeden. Hiermee wordt het belang van de rol van de interviewer in de interactie met de respondent bevestigd en wordt ook duidelijk dat het interviewergedrag nog verder geoptimaliseerd kan worden.

#### 5. De procesevaluatie van de geregistreerde antwoorden

De training van de interviewers en voornamelijk de wijze waarop de interviewers de inhoud van de training in de praktijk omzetten staat centraal in de procesevaluatie van de geregistreerde antwoorden. 'Virtually no one planning a serious survey would question the value of at least some training for the interviewers' (Fowler & Mangione, 1990). De algemene bedoeling van een interviewertraining bestaat erin om meefouten die zich bij het registreren van de antwoorden kunnen voordoen en waarvoor de interviewers verantwoordelijk zijn, te minimaliseren. Daarom moeten interviewers tijdens een training een gedragsrepertoire worden aangeleerd die hen toelaat om adequaat te handelen binnen verschillende situaties die zich kunnen voordoen in de loop van een interview. Dit gedragsrepertoire kan gestructureerd worden rond drie aandachtspunten: de elementaire interviewregels, het induceren van de rol van betrouwbare respondent door het gebruik van nauwkeurige en systematische instructies en het toedienen van negatieve en positieve feedback en ten derde het reactiegedrag van de interviewer ten aanzien van inadequaat respondentengedrag (Billiet en Loosveldt, 1988). Op die wijze wordt het interviewergedrag gestandaardiseerd met de bedoeling interviewereffecten uit te sluiten of te minimaliseren. Verder moet de interviewer de respondent ook instrueren over hoe hij zijn taak als respondent adequaat en optimaal kan uitvoeren.

Er wordt dus verondersteld dat interviews bij vergelijkbare respondenten op een gelijkaardig wijze verlopen. Men kan dan ook verwachten dat er geen significante verschillen zijn tussen de interviewers wat betreft de gemiddelde duurtijd van een interview. Dit is echter niet het geval (multilevel analyse met geslacht leeftijd en onderwijs als controlevariabele:  $z=5.17$ ,  $p<0.01$ ). De gemiddelde duurtijd van het interview bedraagt één uur en een kwart. Op basis van verschillen tussen de interviewers kan 28.5% van de variantie in duurtijd van het interview verklaard worden. Twee interviewers die samen 60 interviews hebben afgevoerd, hebben elk een gemiddelde duurtijd van minder dan een uur, drie andere interviewers daarentegen die samen slechts 29 interviews hebben afgenomen hebben dan een gemiddel-



de duurtijd van meer dan 100 minuten. Er zijn met andere woorden opmerkelijke verschillen tussen de interviewers wat betreft het tempo waaraan een interview wordt afgenomen. Hiermee hebben we een indirecte indicatie dat er verschillen zijn tussen interviewers in de wijze waarop een interview verloopt.

Om het interactieproces en meer in het bijzonder het interviewergedrag in het kader van een procesevaluatie preciezer te kunnen evalueren, moet men per interviewer beschikken over een aantal op band opgenomen interviews. Bij de aanvang van het survey naar culturele verschuivingen in Vlaanderen 2000 volgden 90 interviewers een interviewtraining. De training werd vijf maal georganiseerd en het aantal deelnemers varieerde tussen de 11 en de 28. Elke training duurde iets meer dan 4 uur. De belangrijkste onderdelen waren: aspecten van datakwaliteit in het APS-survey 1999, toelichting bij de belangrijkste interviewregels, het invullen van de APS-vragenlijst door de interviewers, specifieke aandachtspunten bij de vragenlijst en concrete afspraken in verband met introductiebrieven, contact-en vervangingsprocedure, gebruik contactblad en bandopnames.

De interviewers moesten drie interviews integraal op band opnemen. Om tijdig de interviewers te kunnen bijsturen, dienden alle interviewers binnen de veertien dagen na het volgen van de training hun drie eerste afgewerkte interviews terug te bezorgen. Een daarvan diende op band te worden opgenomen.

Bij het uitvoeren van een procesevaluatie wordt men regelmatig geconfronteerd met de praktisch organisatorische problemen bij een dataverzameling. Ter illustratie daarvan vermelden we enkele concrete feiten. Vijf van de 90 interviewers die de training volgden zijn in feite niet gestart met interviewen. Bij 85 interviewers kon een eerste evaluatie gebeuren maar bij drie was er wel geen bandopname beschikbaar. Slechts iets meer dan de helft van de interviewers (52.9%) respecteert de afspraak dat binnen de 14 dagen na het volgen van de training er drie interviews met één bandopname moet afgewerkt zijn. Van de 40 interviewers die zich niet houden aan deze afspraak zijn er 12 waarbij pas na 1 maand (of later) na het volgen van de training een eerste evaluatie kan gebeuren. Deze problemen maakten reeds van bij de aanvang van het veldwerk duidelijk dat de dataverzameling een moeizaam proces zou worden.

### *5.1 Evaluatie van het interviewergedrag op basis van de eerste bandopname*

Bij alle interviewers waarvoor een eerste bandopname beschikbaar was, zijn dezelfde intervieweracties geëvalueerd. Daarbij werd geopteerd om de interviewergedragingen bij verschillende types van vragen – naar aard en vorm – te beoordelen. Het referentiekader bij de evaluaties zijn de regels en instructies die tijdens de training aan bod kwamen. Elke beluisterde passage werd van commentaar voorzien. Deze commentaar fungeerde meteen als feedback voor de interviewers. Er werd dus voor elke interviewer afzonderlijk een evaluatie gemaakt van een aantal belangrijke aspecten van het interviewergedrag. Het gaat hier om: de wijze waarop de vragen gesteld zijn, het al dan niet steeds opsommen van alle antwoordmogelijkheden, het al dan niet steeds verwijzen naar het gebruik van een antwoordkaart, de wijze waarop instructies en verduidelijkingen aan de respondent worden gegeven en de snelheid waarmee het interview wordt afgehandeld. De evaluatie werd uitgedrukt in een score. Een score '10' is toegekend wanneer de interviewer zich steeds correct heeft gehouden aan de aangeleerde interviewrichtlijnen betreffende de verschillende handelingen. Interviewers die zich niet steeds correct houden aan de interviewrichtlijnen maar het toch

nog behoorlijk doen krijgen een score tussen 7 en 9. Interviewers met een score 6 of lager nemen het helemaal niet zo nauw met de aangeleerde richtlijnen. Tabel 8 geeft een overzicht van de verschillende evaluaties op het niveau van de interviewers.

*Tabel 8: Evaluatie van het effectieve interviewergedrag op basis van een bandopname (N=82)*

<i>Interviewerhandeling</i>	<i>N</i>	<i>%</i>
<i>Stellen van de vragen:</i>		
Steeds precies en volledig (score 10)	20	24.4
Bijna steeds precies en volledig (score 7, 8, 9)	55	67.1
Wel eens sterk afwijkend (score $\leq 6$ )	7	8.5
<i>Opsommen antwoordmogelijkheden:</i>		
Steeds precies en volledig (score 10)	36	43.9
Bijna steeds precies en volledig (score 7, 8, 9)	41	50.0
Wel eens sterk afwijkend (score $\leq 6$ )	5	6.1
<i>Verwijzing gebruik antwoordkaart:</i>		
Altijd (score 10)	62	75.6
Meestal (score 7, 8, 9)	17	20.7
Zelden (score $\leq 6$ )	3	3.7
<i>Geven van instructies:</i>		
Steeds correct (score 10)	63	76.8
Bijna steeds correct (score 7, 8, 9)	15	18.3
Wel eens volledig foutief (score $\leq 6$ )	4	4.9
<i>Geven van verduidelijkingen:</i>		
Steeds correct (score 10)	68	82.9
Bijna steeds correct (score 7, 8, 9)	10	12.2
Wel eens volledig foutief (score $\leq 6$ )	4	4.9
<i>Snelheid afhandeling vragen en antwoorden:</i>		
Correct ritme (score 10)	60	73.2
Iets te snel (score 7, 8, 9)	19	23.2
Veel te snel (score $\leq 6$ )	3	3.7

De conclusie uit tabel 8 is alvast dat de gegeven interviewertraining er niet volledig in slaagt om het interviewergedrag te standaardiseren. Het feit dat er gewerkt wordt met interviewers die regelmatig interviewen en hun eigen stijl hebben, is hier niet vreemd aan. Er is sprake van een differentiatie tussen de interviewers naargelang de verschillende geëvalueerde interviewerhandelingen. Slechts één op vier van de interviewers leest de vragen steeds precies en volledig zoals ze opgenomen zijn in de vragenlijst. Een groepje van 7 interviewers durft van deze richtlijn nogal eens sterk af te wijken. De richtlijn betreffende het steeds precies

en volledig opsommen van alle antwoordmogelijkheden wanneer deze opgenomen zijn in de vraagtekst wordt door iets meer interviewers gerespecteerd alhoewel dit nog geen meerderheid is. Drie op vier van de interviewers verwijst wel steeds expliciet naar het gebruik van een antwoordkaart wanneer dit van toepassing is. Het geven van instructies en verduidelijkingen gebeurt eveneens door een meerderheid correct. Tot slot ligt de snelheid bij de afhandeling van het interview bij een kleine minderheid te hoog. Deze interviewers lezen de vragen veel te snel voor en geven onvoldoende tijd aan de respondenten om tot een antwoord te komen.

Op basis van deze globale evaluatie kan een onderscheid gemaakt worden tussen vier groepen van interviewers. Een groepje van 14 interviewers (17.1 %) behaalt steeds een score van 10 op de verschillende geëvalueerde gedragingen. Deze interviewers passen de aangeleerde interviewrichtlijnen steeds correct toe. Een groepje van vier interviewers daarentegen (4.9 %) neemt het helemaal niet zo nauw met de aangeleerde richtlijnen. Hun gemiddelde score is lager dan 7. Een groep van 35 interviewers (42.7 %) behaalt een score lager dan 9.5 en de overige interviewers (n=29 of 35.4 %) behalen een score tussen de 9.5 en 10.

Deze eerste evaluatie werd, conform de uitgangspunten van integrale kwaliteitszorg, gebruikt als feedback voor de interviewers en indien nodig werden bepaalde interviewers bijgestuurd. Voor twee interviewers betekende dit dat ze in het kader van het onderzoek "Culturele verschuivingen in Vlaanderen 2000" geen interviews meer mochten afnemen. Eén interviewer diende het eerste deel van de interviewtraining met de uiteenzetting van de standaard interviewrichtlijnen opnieuw te volgen. Daarna volgde opnieuw een evaluatie van een bandopname. Deze keer kwam de interviewer beter tegemoet aan de richtlijnen en kreeg dan ook het fiat om verder te werken. Vijf interviewers kregen de opdracht eerst nog een interview met een bandopname te bezorgen vooraleer hun volledige pakket af te werken. Vier interviewers zijn aan deze vraag tegemoet gekomen. In het algemeen hebben zij de feedback betreffende hun effectieve interviewergedrag tijdens de 1<sup>e</sup> bandopname ter harte genomen.

Het evalueren van het interviewergedrag op basis van bandopnames van interviewers is een tijdsintensieve aangelegenheid. Het bijsturen van sommige interviewers vertraagt ook het veldwerk. Toch is een dergelijke procesevaluatie een noodzakelijk onderdeel van de kwaliteitsbewaking bij een survey onderzoek. De evaluatie maakt de interviewer immers duidelijk dat zijn werk degelijk wordt opgevolgd. De evaluatie gebeurt ook bij de aanvang van het veldwerk zodat het werk van de goede interviewers positief kan versterkt worden. Verder is het nog mogelijk om zinvol in te grijpen en bij te sturen bij zij die het minder goed doen. Hoewel hier onmiddellijk moet worden aan toegevoegd dat er werd vastgesteld dat één op vijf van de interviewers de feedback niet afwacht. Ervaren interviewers hebben het blijkbaar ook hier moeilijk om de procedure te volgen.

## 5.2 Een evaluatie op basis van telefonische controles

De wijze waarop interviewers hun taak uitvoeren kan ook geëvalueerd worden op basis van informatie die bekomen wordt door de personen die werden geïnterviewd telefonische opnieuw te contacteren en enkele vragen te stellen. De vragen tijdens het telefonisch interview kunnen zowel betrekking hebben op het verloop van het face-to-face interview (o.a. duur van het interview, wie vulde de vragenlijst in en werden de antwoordkaarten gebruikt) als op de herbevraging van een aantal feitelijke gegevens (o.a. opleidingsniveau, werksitua-

tie en familiale situatie). Het is duidelijk dat een dergelijk interview in tegenstelling tot het beluisteren van de bandopname van één van de drie eerste interviews een meer controlerende dan een evaluerende functie heeft. Het werk van de interviewer is immers op het ogenblik van de telefonische herbevraging reeds afgerond zodat er niet meer kan worden bijgestuurd. Bij onregelmatigheden kan men de interviewer ermee confronteren en om de nodig uitleg vragen. In het ergste geval worden de interviews geweigerd en kan de samenwerking met de interviewer worden stop gezet.

Voor het survey naar de culturele verschuivingen in Vlaanderen 2000 werden 27% van de face-to-face bevraagde respondenten (361/1345) telefonisch opnieuw geïnterviewd. Bij 11,9% van de telefonische interviews was de conclusie dat het interview niet volledig volgens de richtlijnen van de training verlopen was (o.a. de respondent zelf de vragenlijst laten invullen en geen antwoordkaarten gebruiken). Daarbij waren 28 interviewers betrokken. Hoewel het zeker niet altijd om grove schendingen van de regels gaat, is het terug een indicatie dat sommige interviewers een eigen 'interviewstijl' hebben en zich niet gemakkelijk schikken naar bepaalde regels.

De vergelijking tussen de telefonisch interviews en de face-to-face interviews voor een aantal feitelijke gegevens is weergegeven in onderstaande tabel. De vastgestelde verschillen bij deze vergelijking zijn opmerkelijk en doen in ieder geval vragen rijzen bij de betrouwbaarheid van deze data.

Tabel 9: Verschillen tussen telefonische controle en resultaten van de face-to-face interviews

	<i>Respondenten waarbij geen overeenstemming is tussen telefoon en face-to-face</i>		<i>Interviewers N=80</i>
	n	%	%
Hoogst behaald diploma	136/354	38.4	73.7
Betaald werk	10/354	2.8	10.0
Beroeps categorie	33/186 *	17.7	29.1
Activiteit geen betaald werk	22/158 *	13.9	22.5
Familiale situatie	14/354	4.0	13.7
Aantal gezinsleden	29/353	8.2	27.5
Onbetaald vrijwilligerswerk	40/354	11.3	40.0

\* Totaal is aantal correcte antwoorden bij betaald werk (n=344)

Bij de vraag naar het hoogst behaald diploma werden heel wat verschillen genoteerd tussen het antwoord bekomen via de telefonische controle en het antwoord op de vragenlijst die tijdens het face-to-face interview werd ingevuld. Ook bij de andere feitelijke vragen is het percentage met verschillen niet gering. Op basis van deze vaststellingen stelt zich uiteraard de vraag welk antwoord als het "correcte" wordt beschouwd wanneer er verschillen optreden. Wat betreft opleidingsniveau kan opgemerkt worden dat in de meeste gevallen het aangrenzende categorieën betreft. Dit ligt in de lijn van de mogelijke verklaring van de ondervertegenwoordiging van lager opgeleiden die werd vastgesteld bij de outputevaluatie van de

gerealiseerde steekproef. Er moet ook nog opgemerkt worden dat tijdens het persoonlijke interview de respondent gebruik kon maken van antwoordkaarten wat het beantwoorden van de vragen gemakkelijker zou moeten maken.

Het is niet altijd duidelijk hoe de resultaten van evaluaties op basis van telefonische interviews gebruikt kunnen worden. Meestal blijft het onduidelijk wat het correcte antwoord is. Interviewers hebben ook meestal een uitleg waarom bepaalde regels niet werden gevolgd en waarom de antwoorden verschillen. De resultaten van de vergelijking zijn doorgaans ook pas beschikbaar wanneer het veldwerk reeds ver gevorderd is. Hierdoor wordt het moeilijk om bij te sturen of om op te treden.

Telefonische controle van de geslaagde persoonlijke interviews blijft echter een belangrijk instrument om het werk van de interviewers te controleren en te evalueren. De telefonische herbevraging is een functioneel onderdeel van de kwaliteitsbewaking van een survey onderzoek. Het verdient aanbeveling om dit in de toekomst te blijven doen. Interviewers die helemaal geen rekening houden met de richtlijnen van de training en waarbij meerdere en frappante verschillen vastgesteld worden tussen de antwoorden op feitelijke vragen in beide interviewsituaties kunnen wel degelijk gedetecteerd worden. Interviewers moeten ook weten dat hun werk gecontroleerd en geëvalueerd wordt. Ook naar de respondenten toe is het een garantie dat kwaliteit wordt nagestreefd. De resultaten komen wel vrij laat. De telefonische interviews zouden bijgevolg vroeger moeten starten en de resultaten ervan moeten systematische worden bijgehouden op het niveau van de interviewers zodat tijdig kan worden bijgestuurd.

De vastgestelde verschillen tussen de telefonische bevraging en het face-to-face interview maken ook bijzonder duidelijk dat het bevragen van op eerste zicht eenvoudige feitelijke kenmerken geen sinecure is en dat ook dit met de grootste zorg dient te gebeuren. Wellicht kunnen een aantal problemen worden opgelost door bij de vraagformulering de gebruikte begrippen preciezer te definiëren en door nauwkeuriger instructies te voorzien voor de interviewers.

### *5.3 Conclusie*

De procesevaluatie van de geregistreerde antwoorden maakt duidelijk dat interviewers zich niet altijd aan afspraken houden en dat daardoor heel wat organisatorische problemen optreden bij een dataverzameling door middel van survey interviews. Er werden ook grote verschillen genoteerd wat betreft de snelheid waarmee de interviews werden afgenomen. Zowel de interactie-analyse op basis van op band opgenomen interviews als de informatie die bekomen werd door middel van de telefonische bevraging maakt duidelijk dat de interviewertraining er niet volledig in slaagt om het interviewgedrag te standaardiseren en dat sommige interviewers belangrijke regels niet toepassen. Dit bevestigt wat reeds bij de outputevaluatie van de geregistreerde antwoorden werd gesteld. Opmerkelijk zijn ook de vastgestelde verschillen tussen wat genoteerd werd tijdens het face-to-face interview en wat geregistreerd werd tijdens het telefonisch opnieuw bevragen van dezelfde respondenten.

## 6. Besluit

Het kader waarin een proces- en een outputevaluatie wordt gekoppeld aan de ideeën van de integrale kwaliteitszorg biedt heel wat mogelijkheden bij de kwaliteitsevaluatie van surveys. In deze bijdrage werden er een aantal van uitgewerkt en geïllustreerd. De concrete toepassing van deze mogelijkheden impliceert in een aantal gevallen wel dat er bijkomend in de dataverzameling geïnvesteerd wordt (o.a. het gebruik van contactbladen, het rapport van de interviewer op het eind van het interview, het laten opnemen op band van interviews, ...). Vooral het uitvoeren van de procesevaluatie vergt heel wat bijkomend werk. Een procesevaluatie levert echter wel belangrijke informatie op over hoe er aan kwaliteitsverbetering kan gewerkt worden. Zo werd bij de outputevaluatie van de geregistreerde antwoorden vastgesteld dat interviewers weten hoe een vragenlijst correct moet worden ingevuld. De procesevaluatie leert ons echter dat het afnemen van de interviews niet altijd volgens de vooropgestelde regels gebeurt. Er moet dus een onderscheid gemaakt worden tussen het correct invullen van een vragenlijst en het volgens bepaalde regels afnemen van het interview. Dit laatste kan ongetwijfeld verbeterd worden. Ook bij de procesevaluatie van de gerealiseerde steekproef werd duidelijk dat het contacteren en vervangen van steekproefelementen niet altijd volgens de vooropgestelde procedures is verlopen. Deze informatie is bijzonder bruikbaar om na te gaan hoe de non-respons fout die werd vastgesteld bij de outputevaluatie kan gereduceerd worden. Procesevaluatie en outputevaluatie leveren dus complementaire informatie op die – conform de ideeën van integrale kwaliteitszorg – kan aangewend worden om de kwaliteit van de verzamelde data te verbeteren. Het is duidelijk dat interviewers in dit proces van kwaliteitsverbetering een belangrijke rol spelen. Het is echter de verantwoordelijkheid van onderzoekers om ze beter op hun taak voor te bereiden en ze beter te ondersteunen bij de uitvoering ervan. Uitgebreidere trainingsprogramma's voor interviewers en een grotere (financiële) waardering voor hun werk zijn ongetwijfeld belangrijke elementen om bepaalde van de vastgestelde problemen te remediëren en te komen tot survey data van hoge kwaliteit.

## Bibliografie

- American Association For Public Opinion Research (2000) *Standaard definitions. Final Dispositions of case codes and outcome rates for surveys*. Ann Arbor: AAPOR
- Billiet, J. & G. Loosveldt (1988) *Improvement of the quality of responses to factual survey questions by interviewer training*. Public Opinion Quarterly, 52, 2, 190-211
- Carton, A. (1999) *Selectie, training en evaluatie van interviewers binnen een interviewer-netwerk*. Leuven: Garant.
- Carton, A., Loosveldt, G. & J. Billiet (1998) *Selectie van interviewers en datakwaliteit in het survey-onderzoek*. Economisch en Sociaal Tijdschrift, 52, 2, 403-427.
- Carton, A. & G. Loosveldt (2000) *Basisdocumentatie: culturele verschuivingen in Vlaanderen 2000*. Leuven: Afdeling Dataverzameling en Analyse, Departement Sociologie, K.U.Leuven.
- De Heer, W. (1999) *International response trends: results of an international survey*. Journal of Official Statistics, 15, 129-142.
- De Leeuw, E. & W. de Heer (1999) *Trends in household survey nonresponse, a longitudinal and international comparison*.. Paper presented at the international conference on survey non-response, Portland, Oregon, USA, 1999
- De Wit, H. (1994) *Cijfers en hun achterliggende realiteit. De MTMM- Kwaliteitsparameters op hun kwaliteit onderzocht*. Leuven: Departement Sociologie, K.U.Leuven
- Fowler, F. & T. Mangione (1990) *Standardized survey interviewing: minimizing interviewer-related error*. Newbury Park: Sage Publications.
- Groves, R. (1989) *Survey errors and survey costs*. New York.: John Wiley.
- Groves, R. & M. Couper (1989) *Nonresponse in household interview surveys*. New York: John Wiley.
- Loosveldt, G. & A. Carton (1997) *Evaluatie van datakwaliteit in surveyonderzoek: het gebruik van het na-interview en het rapport van de interviewer*. Mens en Maatschappij, 72, 2, 166-175.
- Loosveldt, G. & A. Carton (2001) *'Knowledge from Whom ?' Het effect van utilitair individualisme en politieke machteloosheid op de non-respons van een tweede golf van een panel survey*. Mens en Maatschappij, 76, 1, 44-56
- Loosveldt, G. (1985) *De effecten van een interviewtraining op de kwaliteit van gegevens bekomen via het survey-interview*. Leuven: Departement Sociologie, K.U.Leuven.

Pickery, J. (2000) *Applications of Multilevel Analysis in Survey Data Quality Research. Random Coefficient Models for Respondent and Interviewer Effects*. Leuven: departement Sociologie, K.U.Leuven

Purdon, S. , Campanelli, P. & P. Sturgis (1997) *Studying interviewer's calling strategies as a way to improve response rates*. in Campanelli, P. et al. (eds.) *Can You hear me knocking: An investigation into the impact of interviewers on survey response rates*. London: Survey Methods Centre at SCPR.

Stoop, I. & F. Louwen (2000) *'How different are nonrespondents?'* Paper presented at the Fith International Conference on Logic and Methodology, Cologne, October 2000.