



07:26

Henning Olsen

# KONSTRUKTION OG KVALITETSSIKRING AF MULTISURVEYDATA



07:26

# KONSTRUKTION OG KVALITETSSIKRING AF MULTISURVEYDATA

Henning Olsen

KØBENHAVN 2007  
SFI – DET NATIONALE FORSKNINGSCENTER FOR VELFÆRD

## KONSTRUKTION OG KVALITETSSIKRING AF MULTISURVEYDATA

Afdelingsleder: Ole Gregersen

Afdelingen for Socialpolitik og velfærdsydelse

ISSN: 1396-1810

ISBN: 978-87-7487-871-1

Layout: Hedda Bank

© 2007 SFI – Det Nationale Forskningscenter for Velfærd

SFI – Det Nationale Forskningscenter for Velfærd

Herluf Trolles Gade 11

1052 København K

Tlf. 33 48 08 00

[sfi@sfi.dk](mailto:sfi@sfi.dk)

[www.sfi.dk](http://www.sfi.dk)

SFI's publikationer kan frit citeres med tydelig angivelse af kilden. Skrifter, der omtaler, anmelder, henviser til eller gengiver SFI's publikationer, bedes sendt til centret.

# INDHOLD

	FORORD	5
	RESUMÉ	7
	Baggrund	7
	Fra validitet og reliabilitet til ækvivalens	8
	konstruktion af Spørgeskemaer	9
	Respondenters opgaveløsninger	9
	Afprøvning af spørgeskemaer	10
	Behov for multisurveysk metodeudvikling	11
1	BAGGRUND, PROBLEMSTILLINGER OG PROCEDURER	13
	1.1. Indledning	13
	1.2. Baggrund	14
	1.3. Tema, problemstillinger og procedurer	22
2	FRA VALIDITET OG RELIABILITET TIL ÆKVIVALENS	27
	2.1. Indledning	27
	2.2. Monosurveysk validitet og reliabilitet	28
	2.3. Multisurveysk ækvivalens	31

3	SPØRGESKEMAKONSTRUKTION	43
	3.1. Indledning	43
	3.2. Monosurveyske anbefalinger	44
	3.3. Multisurveyske anbefalinger	50
	3.4. Oversættelse	58
4	RESPONDENTERS OPGAVELØSNINGER	75
	4.1. Indledning	75
	4.2. Spørgsmålsforståelse	77
	4.3. Genkaldelse af informationer	92
	4.4. Konteksteffekter	102
	4.5. Svarredigering	108
5	AFPRØVNING AF SPØRGESKEMAER	115
	5.1. Indledning	115
	5.2. Prøveundersøgelser	117
	5.3. Fokusgrupper	119
	5.4. Kognitive laboratorieinterview	121
	5.5. Interaktionskodning	125
	5.6. <i>Split sample</i> -eksperimenter	127
	5.7. Andre procedurer	129
6	SAMMENFATNING	133
	6.1. Indledning	133
	6.2. Problematiserende sammenfatning	137
	LITTERATUR	146
	SFI-RAPPORTER SIDEN 2006	177

# FORORD

Denne rapport handler om, hvordan multisurveydata – dvs. data over landegrænser og på tværs af kulturer – konstrueres og kvalitetssikres. I rapporten besvares følgende problemstillinger: Hvad er måleproblemer i multisurveys, og hvordan begrebsliggøres de? Hvordan formuleres og oversættes spørgsmål over grænser og kulturer? Hvordan løser kulturelt forskellige respondenter deres svaropgaver? Hvordan afprøves multisurveyiske spørgeskemaer?

Rapporten er baseret på metodelitteratur om multisurveys, hvis tendenser kondenseres og sammenholdes med tendenser i metodelitteratur om monosurveys, dvs. enkeltstående surveyundersøgelser.

Litteraturstudiet, hvis udarbejdelse er finansieret af SFI, er gennemført af adj. docent *Henning Olsen*, der takker for konstruktive kommentarer til manuskriptudkast fra seniorforsker *Torben Fridberg* (SFI) og konsulent *Jan Katlev* (Dansk Sprognævn). Endvidere takkes lektor *Christian Albrecht Larsen* (AaU), der har virket som ekstern referee.

København, august 2007

*Jørgen Søndergaard*





# RESUMÉ

Temaet for denne rapport er, hvordan surveydata konstrueres og kvalitetssikres i multisurveys, dvs. surveyundersøgelser over landegrænser og på tværs af kulturer. Rapporten er bygget op om fire problemstillinger:

- Hvordan konceptualiseres måleproblemer i multisurveys (kapitel 2)?
- Hvordan formuleres og oversættes spørgsmål over grænser eller på tværs af kulturer (kapitel 3)?
- Hvordan løser respondenter semantiske og andre kognitive svaropgaver over grænser eller på tværs af kulturer (kapitel 4)?
- Hvordan kan spørgeskemaer afprøves til forebyggelse af multisurveyiske måleproblemer (kapitel 5)?

Rapporten, der er et litteraturstudie, er baseret på multisurveyisk metode-litteratur, hvis tendenser kondenseres og sammenholdes med metodelit-teratur vedrørende monosurveys, dvs. enkeltstående surveys.

## BAGGRUND

Det seneste årtier har været kendetegnet ved en vækst i antallet af multi-surveys. Multisurveys adskiller sig på flere måder fra monosurveys. Mul-

tisurveys fokuserer i særlig grad på ligheder og forskelle imellem lande og kulturer, og det tværnationale eller tværkulturelle perspektiv rejser særlige problemer i forhold til datasammenlignelighed.

Eftersom ikke alle begreber er universelle på tværs af grænser eller kulturer, er det afgørende at afklare, om begreber er kulturuniverselle eller kulturspecifikke. I angelsaksisk litteratur betegnes det universelle perspektiv *etic* og det specifikke *emic*. Mens den 'emiske' tilgang er kulturspecifik, er den 'etiske' tilgang tværgående. I multisurveys bør tilgangen helt eller overvejende være 'etisk', men den viser sig undertiden at være påtvunget 'etisk'.

Derudover er *måden*, respondenter med forskellig kulturbaggrund og nationalitet løser svaropgaver på, afgørende for sammenligninger. Det er en betydelig udfordring i multisurveys, at sprog, sociale konventioner, kognitive evner og svarprocessuelle løsninger veksler kulturer imellem. Forskellene kan let generere data, der peger på kulturelle forskelle, men som faktisk skyldes forståelsesforskelle.

#### FRA VALIDITET OG RELIABILITET TIL ÆKVIVALENS

Den første problemstilling angår i sidste instans, om multi- og monosurveys er grundlæggende forskellige eller ej. Bortset fra at multisurveys overskrider grænser og stiller særlige praktiske krav, er der næppe basale forskelle mellem multi- og monosurveys. I princippet er multisurveys blot flere surveyundersøgelser gennemført i forskellige lande og/eller kulturer. Ganske vist er komparativ fokusering, sammenlignelighed, definitioner af begreber og respondenternes vekslende opgaveløsninger særligt vigtige aspekter i multisurveys, men forekommer også i monosurveys. Hvis der kun er gradforskelle mellem multi- og monosurveys, er det bemærkelsesværdigt, at deres begrebsliggørelse af kvalitetssikring er forskellig. I litteratur om monosurveys er der konsensus om at anvende begreberne validitet og reliabilitet til at forebygge måleproblemer, mens der i litteratur om multisurveys er enighed om at anvende begrebet ækvivalens. Da multisurveys er 'grænseoverskridende', er det oplagt i særlig grad at fokusere på datasammenlignelighed, men det forklarer ikke, hvorfor der skulle være behov for et særskilt kvalitetsbegreb ved siden af

reliabilitet og validitet, idet målinger, der både er pålidelige og gyldige, også burde være sammenlignelige over grænser.

## KONSTRUKTION AF SPØRGESKEMAER

Den anden problemstilling er, hvordan spørgsmål bør formuleres og oversættes. Mens litteratur om monosurveys er rig på forslag til, hvordan spørgsmål kan formuleres og svarkategorier udarbejdes, er langt de fleste anbefalinger i litteratur om multisurveys overtaget direkte fra den monosurveyiske tradition. Det gælder fx spørgsmålsfokusering, længde, sproglig enkelhed og anvendelse af præciserende *probes*.

Bortset fra overvejelserne om det 'etiske' og 'emiske' perspektiv afviger den multisurveyiske anvendelse af begreber ikke fra anbefalingerne i monosurveyisk litteratur. Også monosurveys forudsætter konceptuel sammenlignelighed, ligesom den triangulerende *more is better*-anbefaling af flere indikatorer på samme begreb er velkendt. Når landegrænser eller kulturer overskrides, øges risikoen for, at abstrakte begreber ikke fører til ækvivalente målinger.

Metodelitteraturens helt særlige og omfattende bidrag handler om oversættelse af spørgeskemaer. Selv om litteraturen, der kun er baseret på få empiriske undersøgelser, langt fra er konsensuel, indeholder den adskillige refleksioner og anbefalinger, der ikke blot er formålstjenlige i multisurveys, men som også kan anvendes, når spørgeskemaer i monosurveys skal oversættes. En række specifikke anbefalinger – fx om oversættelseskvalifikationer og samarbejde – er ligeledes nyttige.

## RESPONDENTERS OPGAVELØSNINGER

Den tredje problemstilling er, hvordan respondenter løser kognitive opgaver, og hvordan sammenlignelighed sikres. Til opgaverne hører forståelse af spørgsmål, genkaldelse af informationer, konteksteffekter og redigering af informationer.

Måleproblemer i forbindelse med forståelse af spørgsmål og svarkategorier er grundigt behandlet i metodeforskningen inden for monosurveys, men mere sporadisk inden for multisurveys. Mens monosurveyisk forskning omhandler en lang række problematikker fra forståelse

af ord over spørgsmålssemantik til adskillige svarkategorielle problematikker, er multisurveysk metodeforskning mindre omfangsrig og tematisk snævrere. Det gælder også forskningsbaserede råd om, hvordan man reducerer ækvivalensproblemer. Flere undersøgelser peger dog på, at variationer i 'grænseoverskridende' forståelse af spørgsmål og svarkategorier skaber ækvivalensproblemer. Intet tyder imidlertid på, at der tale om en basalt anderledes metodeproblematik.

Med hensyn til respondenteres genkaldelse af informationer er monosurveysk metodeforskning ligeledes omfattende og bredspektret, mens multisurveysk er sporadisk. Kun om strategier for genkaldelse af informationer samt om kognitiv magelighed findes enkelte resultater. Hvad kognitiv magelighed angår, drejer det sig især om føjelighed (*acquiescence*), der synes mere udbredt i kollektivistiske end individualistiske kulturer. Føjelighed er beslægtet med kognitiv magelighed, fordi begge er kendetegnet ved stillen-tilfreds-strategier (*satisficing*). Alt i alt forbliver det derfor uafklaret, hvornår og under hvilke betingelser genkaldelse er kulturelt sensitiv i multisurveys.

Når spørgsmålsforståelse eller andre svarprocessuelle opgaver afhænger af tidligere spørgsmål, kan målinger blive behæftet med kontekstuelle fejl. Herom findes en omfattende monosurvey-litteratur, der dog ikke udpeger sikre procedurer til forebyggelse af den slags effekter. På den anden side er tilsvarende multisurvey-forskning stort set fraværende. Når konteksteffekter genererer måleproblemer i monosurveys, bør antagelsen være, at effekterne i al fald ikke er af mindre omfang, når grænser eller kulturer overskrides. Der er derfor også et betydeligt forskningsbehov med hensyn til forebyggelse af kontekstuelle ækvivalensproblemer.

## AFPRØVNING AF SPØRGESKEMAER

Den sidste problemstilling er, hvordan multisurveyske måleproblemer kan forebygges. Monosurveys skal honorere krav om validitet og reliabilitet, mens multisurveys skal honorere krav om ækvivalens – et begreb, der kun er vagt defineret, og som findes i over 50 specifikke former. Hvorfor afprøvning af validitet og reliabilitet ikke er tilstrækkelig, besvarer multisurveysk litteratur ikke.

Yderligere en væsentlig forskel er mangel på undersøgelser af, hvordan prætest virker, og hvor effektive de er. Indtil dette er blevet undersøgt, foreslår toneangivende forskere nogle almene anbefalinger, der er blevet sammenfattet af Smith (i: Presser et al. 2004). Anbefalingerne er rettet mod tværnationale surveys, men kan også overføres til tværkulturelle:

1) the multipretesting procedures should be carried out across countries and languages with the results evaluated by researchers expert in (a) the cultures and languages being investigated, (b) the substantive domains being studied, and (c) survey research methodology; (2) the pretesting and translation should be integrated and interactive processes; (3) the pretesting needs to be cross-national and not just within specific countries; and (4) the developmental process takes much more time and resources than for single-country monolingual studies and usually should involve multiple rounds of pretesting and larger samples.

## BEHOV FOR MULTISURVEYSK METODEUDVIKLING

I rapportens afsluttende kapitel peges på et behov for yderligere multi-survey-sk metodeudvikling inden for konstruktion og afprøvning af ”grænseoverskridende” spørgeskemaer, fx komparative analyser af kvalitetsstyrings begrebslige grundlag, analyser af multi-survey-sk spørgeskemakonstruktionens eventuelle specifikke problemer, undersøgelser af, om respondenters opgaveløsninger skaber større måleproblemer i multi- end i monosurveys, undersøgelser af multi-survey-ske konteksteffekter samt undersøgelser af fordele og ulemper ved forskellige måder, hvorpå multi-survey-ske spørgeskemaer kan afprøves.

Som ofte påpeget i metodelitteratur om multi-surveys er der derudover behov for et øget tværfagligt samarbejde, fx inspireret af CASM-bevægelsen (*Cognitive Aspects of Survey Methodology*), der i de senere år har ydet betydelige bidrag til metodeudvikling inden for monosurveyområdet. Det drejer sig fx om samarbejde med lingvister, fx socio-semantikere, men flere multi-survey-ske forskere mener også, at kognitionspsykologer, kommunikationsforskere m.fl. kan bidrage til metodeudvikling.



# BAGGRUND, PROBLEMSTILLINGER OG PROCEDURER

## 1.1. INDLEDNING

Denne rapport, der henvender sig til forskere og studerende inden for sociologi og tilgrænsende samfundsvidenskabelige discipliner, sætter fokus på, hvordan surveydata konstrueres og kvalitetssikres med særlig henblik på data over landegrænser og på tværs af kulturer.

Der redegøres først summarisk for faser, som monosurveys – dvs. enkeltstående surveyundersøgelser – almindeligvis omfatter, hvorefter der introduceres til multisurveys, dvs. surveyundersøgelser på tværs af landegrænser og kulturer (afsnit 1.2).<sup>1</sup> Med afsæt heri præsenteres rapportens tema, problemstillinger og procedurer (afsnit 1.3). Det nævnte tema er stort set upåagtet i dansk kvantitativ metodelitteratur og dermed én begrundelse for rapporten, der bl.a. kan anvendes ved vurdering af resultater fra et stadigt voksende antal multisurveys.<sup>2</sup>

- 
1. Der findes også andre typer af multisurveys, fx nationale longitudinelle, der imidlertid ekskluderes af rapporten. Det var oprindelig hensigten tillige at sætte fokus på konstruktion og kvalitets sikring af surveydata i multisurveys gennemført over tid, men en præliminær litteratursøndering afdækkede så få relevante videnskabelige bøger og artikler, at et anvendelsesorienteret litteraturstudie ikke er gennemførligt.
  2. Ordene multi- og monosurveys er valgt for at undgå lange betegnelser, fx ”tværnationale og tværkulturelle surveyundersøgelser”.

Rapportens tema udmøntes i fire problemstillinger, der belyses i et tilsvarende antal kapitler. Den første problemstilling omhandler, hvad måleproblemer vil sige multisurveysk, og hvordan de begrebsliggøres (kapitel 2), mens den anden sætter fokus på formulering og oversættelse af spørgsmål over grænser og på tværs af kulturer (kapitel 3). Den tredje problemstilling angår respondenters opgaveløsninger og løsningernes konsekvenser for udarbejdelse af multisurveyske spørgeskemaer (kapitel 4), mens den sidste omhandler afprøvning af spørgeskemaer (kapitel 5).

Rapporten baseres på multisurveysk metodelitteratur, hvis tendenser kondenseres og sammenholdes med tendenser i monosurveysk litteratur, der foreligger kondenseret i bogen *Fra spørgsmål til svar* (Olsen, 2005). Rapporten er derfor et litteraturstudie, der har forudsat søgning, tilvejebringelse, læsning og kondensering af et bredspektret udvalg af multisurveysk metodelitteratur. Mens kapitel 2-5 er en tilstræbt loyal kondensering af mono- og multisurveysk – navnlig engelsksproget – metodelitteratur, udgør rapportens afsluttende kapitel en problematiserende sammenfatning, hvor rapportens forfatter kommer direkte til orde og påpeger et markant behov for multisurveysk metodeudvikling.

## 1.2. BAGGRUND

Der redegøres først for rapportens baggrund, idet der sættes fokus på faser, som samfundsvidenskabelige monosurveys normalt omfatter, hvorpå der introduceres til multisurveyske problematikker.<sup>3</sup>

Blandt kvantitative metoder – fx surveys og registerundersøgelser – har førstnævnte stor udbredelse, fx i de nordiske og angelsaksiske lande (se fx Andersson, 1994; Olsen, 2001a). I metodelitteraturen er der imidlertid ikke konsensus om, hvordan surveymetode bør defineres (se fx Fink & Kosecoff, 1998; Folz, 1996; Rea & Parker, 1997). Surveys påpeges dog ofte at være undersøgelser, hvor data tilvirkes ved hjælp af spørgeskemaer, hvis spørgsmål stilles mundtligt til en stikprøve af respondenter eller som selvudfyldes (se fx Aldridge & Levine, 2001), mens andre i højere grad betoner metodens deskriptive potentiale (se fx Sapsford, 1999).

---

3. Da de fleste læsere forventeligt har kendskab til monosurveys, er beskrivelsen af monosurveys summarisk (uddybende, se fx Olsen, 2005).



Surveydata, der normalt tilvirkes ved tilstræbt standardiserede interview, er i princippet egnede til repræsentative kortlægninger af udbredelsen af og sammenhænge mellem klassificerbare sociale eller andre fænomener, selv om måleproblemer i praksis kan blokere herfor, fx på grund af stikprøveproblemer, mangelfuld spørgeskemakonstruktion og respondenteres uvilje mod at lade sig interviewe (se fx Fink & Kosecoff, 1998). Der er derfor "... considerable potential for error in the design and execution of surveys" (Kent, 2001:10). Surveyresultater opnås ved at addere respondenteres svar til aggregerede indikatorer for fænomeners udbredelse og statistiske sammenhænge.

Trods manglende konsensus angående surveydesigns (se fx Folz, 1996; Jackson & Furnham, 2000; Sapsford, 1999), gennemløber de normalt en række faser: forberedelsesfase med tematisering og design, populations- og stikprøvefase, spørgeskema- og interviewfase, dataoparbejdsfase samt analyse- og rapporteringsfase. Under tematiseringen afgøres fx, om en survey relateres til teoretisk eller anden forforståelse. Selv om surveys ofte er deskriptive, muliggør de fx også teoritest (se fx Aldridge & Levine, 2001; de Vaus, 1998). At formulere teoriirrelaterede, deskriptive eller andre problemstillinger er at udarbejde, afgrænse og præcisere de forskningsspørgsmål, som en surveyundersøgelse sigter mod at belyse.

Med udgangspunkt i en præcis og velafgrænset population udtrækkes en stikprøve med henblik på at drage slutninger fra stikprøve til population (se fx Aldridge & Levine, 2001; Fink, 1995a; Fowler, 1993; Harpaz, 2000). Surveymetoden er med andre ord induktiv, eftersom slutninger drages fra "nogle" til "alle". I forlængelse heraf er nogle vigtige spørgsmål: Hvordan afgrænses populationen? Kan der opnås adgang til den? Efter hvilke kriterier skal stikprøven udtrækkes og af hvilken størrelse? Udtrækningen er som regel sandsynlighedsbaseret, da induktive slutninger ellers er umulige. Bestræbelsen er, at alle besvarer spørgeskemaet. Der forekommer imidlertid altid bortfald, fx på grund af uvilje mod at deltage i undersøgelsen, hvilket også kan føre til måleproblemer (se fx Fowler, 1993; Groves, 2001; Groves, 1989). Hvis bortfaldet er tilfældigt, er analytiske reservationer unødvendige. Det modsatte forekommer dog i næsten alle tilfælde.

Konstruktion og besvarelse af spørgeskemaer er en af surveymetodens skrøbeligste aspekter (se fx Alwin, 1991). Problemstillinger indeholder ofte begreber, som respondenter intet hører om. Da respondenter er datatilvirkningens medier, ændres abstrakte begreber til hverdagsprog

og fører respondenter ad et spor, hvor mening er skabt på forhånd (se fx Fowler, 2002): Hvordan defineres og operationaliseres begreber i form af spørgsmål, som respondenter kan forstå og genkalde informationer i relation til?

Hvis besvarelse af spørgsmål ikke skal skabe måleproblemer, skal spørgsmål formuleres så præcist som muligt. Kravet er trivielt, men hvordan honoreres det i praksis? Ofte anvendes "tommelfingerregler", der dog ikke er garant for uproblematisk målinger. Forskere har derfor udviklet forskningsbaserede procedurer til fremme af sammenlignelige data (se fx Olsen, 2005). Spørgeskemaer afprøves normalt på "prøverespondenter". Udfaldet af sådanne interview er dog ingen garant for datakvalitet.

Når data tilvirkes af respondenter og interviewere, kan interaktionen skabe måleproblemer (se fx Fowler, 1995; Schuman & Presser, 1996). Surveydata "indsamles" ikke "som blomster på en mark", men konstrueres på måder, der genererer måleproblemer, som statistiske metoder ikke kan neutralisere (se fx Kent, 2001). I så henseende er nogle centrale spørgsmål: Forstår respondenter spørgsmål som tilsigtet? Forstår spørgsmål på samme måde af alle? Smitter forståelse af ét spørgsmål af på forståelsen af de efterfølgende? Muliggør respondenter hukommelse genkaldelse af ønskede informationer? Er respondenter magelige, fx i form af "slag på tasken"? Ved besøgsinterview kan samspillet skabe måleproblemer, fx via *probing*-praksisser, sproglige intonationer og mimik (se fx Fowler & Mangione, 1991; Houtkoop-Steenstra, 2000). Også ved telefoninterview kan interaktion skabe måleproblemer, fx fonetiske. Hvad endelig postspørgeskemaer angår, stiller de undertiden urealistiske krav til læsevne.

De to sidste faser er dataopbejdelse og -analyse (se fx Bryman, 2000; Hellevik, 1993a). Data rejser flere spørgsmål: Hvordan afledes en relevant analysestrategi af problemstillingerne? Hvilke tabeller og figurer skal udarbejdes? Hvilke variabler fungerer som forklarende versus afhængige variabler? Hvilke rekodninger, indeks og/eller skalaer er relevante? Hvilke central- og spredningsmål er relevante? Statistiske test og multivariate modeller hjælper til at drage repræsentative slutninger om populationen (se fx Menard, 1995). Beskrivelsen af tendenser og sammenhænge suppleres med statistiske forklaringer, fx multivariate analyser (se fx Byrne, 2002; Hosmer & Lemeshov, 1989).

Den summariske redegørelse for monosurveys danner baggrund for beskrivelse af komparative surveys på tværs af landegrænser og kulturer. De seneste 30-40 år har været kendetegnet ved en vedvarende vækst i antallet af surveyundersøgelser, der sætter fokus på sammenligning af lande, regioner eller kulturer: "... without cross-cultural comparison, we could not talk or write about what may be universal and variable about human cultures" (Ember & Ember, 2001:2). Blandt andet som følge heraf er der dannet en række forskningsmiljøer med særlig fokus på komparative surveys. Som eksempel kan nævnes *Centre for Survey Research and Methodology* (ZUMA) i Mannheim, hvor bl.a. Janet Harkness og Peter Ph. Mohler yder bidrag til metodeudvikling.<sup>4</sup>

I betragtning af globalisering og øget europæisk integration er væksten ikke overraskende: "... in the context of European integration, reliable cross-national data of a high and consistent quality are not just desirable but essential" (O'Shea et al., 2005:3). Når befolkninger og kulturer kommer tættere på hinanden, øges behovet for sammenlignelige data, fx om befolkningers levevilkår, holdninger og værdier (se fx Mohler, 2005). Det gælder også data, der overskrider Europas ydre grænser, fx i forbindelse med den multi-kontinentale *World Values Survey* (se fx Therborn, 1999). Nogle forskere problematiserer dog multisurveys generelle anvendelighed og bifalder forskellige "kulturelle tilpasninger" (se fx Kuechler, 1998).

Multisurveys sætter fokus på mange temaer, fx økonomiske og sociale indikatorer, holdninger og værdier samt befolkningers evne til at læse og skrive. Blandt eksempler på større tværnationale surveys kan nævnes:

- *Afrobarometer (AF)*
- *Eurobarometer (EB)*
- *European Community Household Panel (ECHP)*
- *European Values Study (EVS)*

---

4. Nogle yderligere eksempler: *Department of Sociology*, University of Leuven (bl.a. Jaak Billiet). *Department of Methodology and Statistics*, Utrecht University (bl.a. Joop J. Hox og Edith D. de Leeuw). *Department of Psychology*, Tilburg University (bl.a. Fons J.R. Van de Vijver). *Department of Statistics and Methodology*, University of Amsterdam (bl.a. Willem E. Saris). *National Opinion Research Centre*, University of Chicago (bl.a. Tom W. Smith). *Survey Research Center at the Institute for Social Research*, University of Michigan (bl.a. Mick P. Couper). *Survey Research Laboratory*, University of Illinois (bl.a. Timothy P. Johnson).

- *European Social Survey (ESS)*
- *EU Statistics on Income and Living Conditions (EU-SILC)*
- *Harmonized European Time Use Survey (HETUS)*
- *International Social Survey Programme (ISSP)*
- *Multinational Time Use Study (MTUS)*
- *Survey of Health, Ageing and Retirement in Europe (SHARE)*
- *World Values Survey (WVS)*

Hertil kommer fx også *Programme for International Student Assessment (PISA)*, *Progress in International Reading Literacy Study (PIRLS)* og *International Adult Literacy Survey (IALS)*. Siden 1970'erne er mange surveyprogrammer designet med specifikt tværnationalt sigte, hvoraf ESS, Eurobarometer, ISSP og WVS hører til de mest kendte.<sup>5</sup>

Blandt langt over 100 ESS-databaserede videnskabelige artikler, papers, arbejdsrapporter mv. skal blot ét eksempel nævnes.<sup>6</sup> Undersøgelsen rejser spørgsmålet, om danskere er specielle, hvad holdninger til immigration angår (Skaksen et al., 2006). Det viser sig, at den typiske dansker er lidt mere skeptisk over for immigration end andre europæere. Desuden adskiller danskerne sig ved, at ret få forbinder øget immigration med negative konsekvenser for arbejdsmarkedet, mens relativt flere forbinder immigration med højere velfærdsstatslige udgifter. Med hensyn til ISSP-data er de ligeledes anvendt som grundlag for udarbejdelse af over 100 artikler, papers mv.<sup>7</sup>

I sociologi og andre samfundsvidenskaber er sammenligninger nødvendige og et integreret aspekt af den sociologiske disciplin fra dens oprindelse. Det gælder fx i Durkheims klassiske undersøgelser af selvmord (se fx Mills et al., 2006). Selv om multisurveys i princippet blot er flere samtidige surveys gennemført i forskellige lande og/eller kulturer, adskiller de sig fra monosurveys på flere måder, fx med hensyn til kom-

---

5. AF: [www.afrobarometer.org/](http://www.afrobarometer.org/), EB: [http://europa.eu.int/comm/public\\_opinion/index\\_en.htm](http://europa.eu.int/comm/public_opinion/index_en.htm), ECHP: <http://epunet.essex.ac.uk>, ESS: <http://www.europeansocialsurvey.org/>, EVS: <http://www.europeanvalues.nl/>, ISSP: <http://www.issp.org/>, MTUS: <http://www.iser.essex.ac.uk/mtus/>, PISA: <http://www.pisa.oecd.org/>, SHARE: <http://www.share-project.org/>, WVS: <http://www.worldvaluessurvey.org/>

6. Yderligere eksempler, se: <http://naticent02.uuhost.uk.uu.net/publicity/index.htm>

7. Se: [www.issp.org/biblio.htm](http://www.issp.org/biblio.htm).

parativ fokusering, datasammenlignelighed, begreber og begrebsdefinitioner samt respondenters vekslende opgaveløsninger.<sup>8</sup>

Multisurveys kendetegnes ved udtalt *komparativ fokusering*, hvis særlige problemer har været kendt længe (se fx Przeworski & Teune, 1970; Øyen, 1990). Det drejer sig både om tværnational og ”tværkulturel” surveyforskning. Når surveys er tværnationale, retter statistiske analyser sig imod ligheder og forskelle lande imellem, fx ved at undersøge "... particular issues of phenomena in two or more countries with the express intention of comparing their manifestations in different socio-cultural settings" (Hantrais & Mangel, i: Hantrais & Mangel, 1996:1). Derimod sætter ”tværkulturelle” surveys i sagens natur fokus på kulturspecifikke ligheder og forskelle.

Nogle forskere anvender dog ”nation” og ”kultur” som synonyme eller ”nation” som ”kultur”-proxy, hvorved der abstraheres fra, at landes befolkninger kan tilhøre forskellige kulturer (se fx Harpaz, 2000). Det er generelt uklart, hvad begrebet ”kultur” betyder, og der er langt fra konsensus herom blandt ”tværkulturelle” forskere (se fx Roney & Taras, 2006; Smith & Bond, 1998). Kulturbegrebets betydning synes at afhænge af den specifikke (teoretiske) sammenhæng, hvori begrebet bruges, fx sociologiske eller antropologiske. Af den grund tjener følgende definition udelukkende som eksempel: ”We interpret culture to represent a social group with a shared language and set of norms, values, beliefs, expectations and life experiences” (Johnson et al., 1997:87).

Måleproblemer er velkendte i surveys: ”With direct measurement we even have problems in national survey research because national societies are different in class affiliations, in education and in cultural history” (Hoffmeyer-Zlotnik & Harkness, 2005:6). Det er efterhånden kendt, at data fra én og samme surveyundersøgelse ikke nødvendigvis honorerer krav om *sammenlignelighed* på tværs af forskellige befolkningsgrupper.<sup>9</sup> Manglende intern sammenlignelighed kan fx gøre sig gældende med hensyn til respondenters køn, alder, etniske tilhørsfor-

---

8. Hertil kommer fx også afgørelse vedrørende deltagende lande og/eller kulturer samt kriterier for stikprøveudtrækning i forskellige lande eller områder (se fx Lynn et al., i: Harkness, 2005).

9. Navnlige amerikanske metodeforskere har i de seneste 10-15 år bidraget til belysning af denne problematik, fx Norman Bradburn, Seymour Sudman, Norbert Schwartz og Roger Tourangeau, der tilhører den såkaldte CASM-bevægelse, dvs. *Cognitive Aspects of Survey Methodology* (se fx litteraturkondenseringen i: Olsen, 2005, kap. 3). I amerikansk metodeforskning anskueliggør mange undersøgelser, der sædvanligvis er *split sample*-baserede, manglende intern sammenlignelighed.

hold, uddannelse og beskæftigelse (se fx Olsen, 2001a). Måleproblemer øver ikke kun indflydelse på gennemsnit og spredning, men kan også påvirke korrelationer. Det bagvedliggende problem er stimulus-respons-modellens næppe opnåelige krav om ”standardiseret” datatilvirkning: ”... even the routinely assumed within-survey functional equivalence may more often be an unproven assumption than an established reality” (Alwin et al., 2002:294).

I multisurveys opstår der imidlertid yderligere komplikationer: “Each culture views life in a unique fashion. This is usually based on the norms, values, attitudes, experiences, etc., particular to that specific culture. Thus, the comparability of any phenomenon could pose a major methodological problem in international research” (Bulmer, 1998:155). Forskellige spørgeskemaer og kulturelt forskellige respondenter kan derfor generere usammenlignelige data (se fx Saris, 1998). Selv om også data fra monosurveys kan være usammenlignelige, formodes problemet at være større i tværnationale og andre multisurveys: “... the opportunities for error are multiplied many times over when one moves from carrying out research in a single society, to doing it in many different societies” (Inglehart & Welzel, 2004:2). Af multisurveys fokus på ligheder og forskelle mellem lande og kulturer følger derfor et særligt fokus på data-sammenlignelighed (se kapitel 2). Herfor taler også det stadigt voksende antal multisurveys.

Yderligere et problem er identifikation af relevante *tværgående begreber* samt definition og operationalisering af disse: “The concepts and items must be relevant cross-nationally and are consequently often not the same ones that would be chosen for a national survey on the same topic” (Lynn et al., i: Harkness, 2005:12). Selv om også monosurveys principielt er komparative, rejser multisurveys særlige problemer: “... cross-cultural survey research is required to pursue strategies that try to come to terms with the fact that concepts may not be identical or comparable and that an instrument appropriate and adequate in one context (temporal or spatial) may not be adequate in another” (Harkness et al., i: Harkness et al., 2003:8).

Da ikke alle begreber er kulturelt universelle, er det nødvendigt at tage stilling til, om multisurveyske problemstillinger er af kulturuniversel eller kulturspecifik beskaffenhed. I angelsaksisk litteratur benævnes det universelle perspektiv almindeligvis *etic*, mens det specifikke kaldes *emic* (i det følgende ”oversat” til ”etisk” henholdsvis ”emisk”, fordi ter-

merne er uoversættelige til dansk).<sup>10</sup> Mens den ”emiske” tilgang fokuserer ”indefra” på specifikke kulturer, har den ”etiske” tilgang bredere komparative synsvinkler og retter sig imod to eller flere forskellige kulturer (se fx Berry, 1999). Ved gennemførelse af multisurveys bør tilgangen helt eller i det mindste overvejende være ”etisk”, men kan vise sig at være påtvungen, hvis ”emiske” forskelle ikke desto mindre gør sig gældende: “... researchers often inappropriately use the etic approach to make cross-cultural comparisons without fully taking into account some relevant culture-specific emics” (Schaffer & Riordan, 2003:174). I angelsaksisk litteratur kaldes dette fænomen ofte *imposed etics* (se fx Johnson, 1998).

Endelig er også kulturelt forskellige respondenters måde at løse *kognitive svarprocessuelle opgaver* på afgørende for, om mulighederne for statistiske sammenligninger vanskeliggøres af måleproblemer: “The great challenge in crossnational survey research is that languages, social conventions, cognitive abilities, and response styles all vary across societies. To obtain valid, equivalent measures across cultures, measurement error from these sources must be minimized and equalized so that valid, reliable, and consistent substantive information emerges. Achieving this is difficult” (Smith, 2004:10). Kulturelt betingede forskelle i måder, hvorpå spørgsmål forstås, kan føre til data, der synes at afdække kulturelle forskelle, men som fx skyldes forståelsesforskelle, forskelle i genkaldelsesstrategier og/eller forskelle i grader af redigering af genkaldte informationer (se fx Johnson et al., i: Hoffmeyer-Zlotnik & Harkness, 2005; Jowell, 1998).

Som følge af ovennævnte og andre særtræk erklærer nogle forskere, at multisurveys forudsætter en helt særlig metodologi: ”... it has been argued that cross-cultural surveys do not require a methodology of their own. This view is, however, not correct” (Mohler, 2005:1). Til støtte for den tilsyneladende ret udbredte opfattelse fremføres, at multisurveys er mere komplekse end monosurveys, fx på grund af flere interviews og mere omfattende datamateriale, ligesom problemer i forbindel-

---

10. Angående *etic* versus *emic*: “The terms originated in linguistic. “Etic” refers to sounds common to all languages and “emic” indicates sound specific to one language or a group of related languages (...). In cross-cultural studies, the etic approach assumes that there is a set of universal cultural dimensions that are equally relevant to all cultures (...). On the contrary, emic approach is based on the assumption that at least some cultural dimensions are culture-specific and cannot be used to analyze cultures of various entities” (Rowney & Taras, 2006).

se med kulturelt tværgående begrebsdefinitioner og begrebsoperationaliseringer nævnes. Det er dog en opfattelse, der kan anfægtes (se kapitel 6).

Ligesom i den omfattende metodelitteratur om monosurveys (se fx Olsen, 2005) giver ovenstående et fingerpeg om, at der næppe heller er konsensus om, hvordan multisurveys bør designes og gennemføres, hvilket da også påpeges af en forsker med betydelig erfaring inden for feltet: "... there is currently little consensus regarding how best to establish cross-cultural equivalence when conducting social surveys" (Johnson, 1998:2). Endvidere synes anbefalinger vedrørende multisurveys – det gælder ikke mindst forskningsbaserede anbefalinger – underudviklede i forhold til monosurveys: "... a calibrated methodology for cross-cultural survey research is not yet available" (Harkness, 1999:135). Der er derfor gode grunde til at udvikle og forfølge problemstillinger, som sætter fokus herpå.

### 1.3. TEMA, PROBLEMSTILLINGER OG PROCEDURER

Ikke mindst på grund af et stadigt voksende antal surveys, der forudsætter sammenlignelige data over grænser eller kulturer, og dansk deltagelse i adskillige af den slags undersøgelser, er der gode grunde til at sætte fokus på måleproblemer af særlig art, som sammenligning af data på tværs af surveys rejser, og på procedurer til forebyggelse heraf. Det blev i afsnit 1.2. anskueliggjort, at mange forhold kan skabe måleproblemer, der anfægter multisurveyske forskningsresultater. I de følgende kapitler sættes der fokus på udvalgte problematikker i tilknytning til spørgeskemakonstruktion.

Rapportens tema er derfor, hvordan surveydata konstrueres og hvordan datakonstruktion kan kvalitetssikres ved gennemførelse af surveyundersøgelser over grænser og på tværs af kulturer. Hermed ekskluderes nationale komparative tværsnitsundersøgelser og nationale longitudinelle undersøgelser af rapporten.<sup>11</sup> Selv i enkeltstående surveys har

---

11. Komparative tværsnitsundersøgelser følger ikke samme respondenter over tid, men forudsætter tidsmæssige sammenligninger, som de fx kendes fra SFI's levekårsundersøgelser, kultur- og fritidsundersøgelser mv. Også nationale longitudinelle undersøgelser forudsætter sammenligninger over tid, fx i SFI's børne- og ældreforløbsundersøgelser. Udelukkelsen af disse undersøgelsestyper



måleproblemer vist sig uundgåelige, men hvilke yderligere måleproblemer kommer på tale, når grænser eller kulturer overskrides, og hvordan lader problemerne sig forebygge? Temaet udmøntes i fire problemstillinger, der belyses i fire kapitler (se *oversigt 1.1.*). Problemstillingerne belyses med afsæt i multisurveysk metodelitteratur, der sammenlignes med tendenser i metodelitteratur vedrørende nationale eller ”monokulturelle” surveys. Da de fleste af rapportens læsere forventeligt har kendskab til monosurveysk metodelitteratur, er kondenseringen heraf summarisk (se Olsen, 2005).

### **Oversigt 1.1.**

#### Rapportens problemstillinger.

---

- Hvad er multisurveyske måleproblemer, og hvordan konceptualiseres problemerne?
  - Hvordan formuleres spørgsmål over grænser eller på tværs af kulturer, og hvordan oversættes spørgsmål i multisurveys til forebyggelse af måleproblemer?
  - Hvordan løser respondenter semantiske og andre kognitive opgaver over grænser og på tværs af kulturer?
  - Hvordan afprøves spørgeskemaer til forebyggelse af multisurveyske måleproblemer?
- 

Den første problemstilling om konkurrerende konceptualiseringer af, hvad måleproblemer er i multisurveysk sammenhæng, er samtidig en belysning af, om der er konceptuelle forskelle mellem multi- og monosurveysk metodelitteratur (kapitel 2). Hvad den anden problemstilling om formulering af spørgsmål angår, retter mange anbefalinger om spørgeskemakonstruktion sig imod alenestående surveys, mens metodelitteraturen synes af begrænset omfang med hensyn til formulering af sammenlignelige spørgsmål over grænser og på tværs af kulturer. Dog forventes metodelitteraturen mere udførlig med hensyn til oversættelse af spørgsmål i multisurveyske spørgeskemaer. Litteraturen omhandler fx oversættelsesprocedurer, hjælpemidler og evaluering af oversættelser (kapitel 3).

Metodelitteratur vedrørende den tredje problemstilling om måleproblemer som følge af nationalt og kulturelt forskellige respondents opgaveløsninger forventes ligeledes begrænset sammenlignet med analyser og anbefalinger vedrørende virkninger af respondents opgaveløsninger i monosurveys, fx spørgsmålsforståelse, kognitiv magelighed og ved ikke-svars omfang. Eksempelvis vil ”samme” spørgsmål gentaget i et

---

per skyldes, at meget få metodeforfattere sætter fokus herpå. Det gælder ikke mindst metodeforskning, der retter sig imod semantiske og andre kognitive ændringer over tid.

andet land eller anden kultur semantisk kunne fungere som et andet spørgsmål, fordi spørgsmåls betydning afhænger af kontekster, hvori de stilles (kapitel 4). Det samme har dog også vist sig i monosurveys.

Hvad endelig den sidste problemstilling om kvalitetssikrende afprøvning af spørgeskemaer angår, forventes tilsvarende forskelle mellem omfanget og arten af mono- henholdsvis multisurveyisk litteratur, fx med hensyn til kognitive laboratorieinterview, interaktionskodning og ekspertvurderinger (kapitel 5). Hertil kommer et afsluttende kapitel, der sammenfatter og problematiserer de foregående kapitler, og hvor rapportens forfatter som nævnt kommer direkte til orde.

Da problemstillingerne belyses med afsæt i tilgængelig metodelitteratur, er rapporten et litteraturstudie, der næsten udelukkende baseres på engelsksproget multisurveyisk metodelitteratur. Litteraturens tendenser kondenseres og sammenholdes med tilsvarende tendenser i monosurveyisk litteratur, der som nævnt allerede foreligger kondenseret (Olsen, 2005). Det har derfor været nødvendigt at søge, fremskaffe, granske og kondensere relevant metodelitteratur. Litteraturen, der omfatter værker såvel som videnskabelige artikler, er søgt på Danbib, Web of Science og Google Scholar ved anvendelse af relevante og trunkerede søgeord:

- compar?
- cross?
- data?
- emic?
- etic?
- equivalent?
- error?
- measurement?
- question?
- reliab?
- survey?
- translat?
- valid?

Efter at litteraturen – næsten 200 bogtitler og artikler – var tilvejebragt af Socialforskningsinstituttets Bibliotek, blev der udarbejdet omfattende læsenoter (cirka 800 sider), som blev kodet og reorganiseret med afsæt i ovennævnte fire problemstillinger. Ved efterfølgende søgning på Google viste det sig, at allerede lokaliseret litteratur undertiden var omtalt dér, fx som kommentarer, pressemeddelelser o.l. I nogle tilfælde er Google-dokumenter derfor umiddelbart overflyttet til læsenoterne.

Bortset fra at den metodelitteratur, som *Fra spørgsmål til svar* baseres på, næsten kun omfatter bogtitler (cirka 180), er søge-, læse- og kondenseringsproceduren i hovedsagen den samme som i denne rapport

(for en nærmere redegørelse, se Olsen, 2005, kap. 1). Da bogen er offentliggjort i 2005, er litteraturreferencerne ikke opdateret og derfor af lidt ældre dato.



# FRA VALIDITET OG RELIABILITET TIL ÆKVIVALENS

## 2.1. INDLEDNING

Hvad enten surveydata hidrører fra mono- eller multisurveys, forudsætter konstruktion og analyser af data empirisk kvalitetssikring. Heraf følger et behov for, at kvalitetssikring konceptualiseres, dvs. at det defineres, hvad et måleproblem er. Som andre begreber har måleproblem ikke på forhånd given betydning og bør derfor defineres. Ellers vil det være umuligt at afgøre, om måleproblemer er store eller ubetydelige. Problemstillingen er derfor:

- Hvad er multisurveyske måleproblemer, og hvordan konceptualiseres problemerne?

Problemstillingen belyses som nævnt med afsæt i et bredspektret udvalg af multisurveysk metodelitteratur, der sammenlignes med litteratur vedrørende monosurveys. Den følgende fremstilling er derfor også en belysning af, om der er afgørende forskelle mellem multi- og monosurveysk metodelitteratur. Kapitlet indledes med en summarisk redegørelse for formodentlig velkendte opfattelser og definitioner af, hvad måleproblemer er, når det handler om den digre monosurveyske metodelitteratur (afsnit 2.2.). Med dette som komparativt ”bagtæppe” sættes der fokus på

ofte forekommende opfattelser og definitioner af måleproblemer i multisurveysk litteratur (afsnit 2.3.). Konceptuelle ligheder og forskelle mellem mono- og multisurveysk litteratur vurderes og problematiseres i rapportens afsluttende kapitel (afsnit 6.2.)

## 2.2. MONOSURVEYSK VALIDITET OG RELIABILITET

Som det vil være de fleste læsere bekendt, er monosurveysk metodelitteratur kendetegnet ved reliabilitet (pålidelighed) og validitet (gyldighed) som bærende begreber med henblik på forebyggelse af måleproblemer. De to begreber virker som gennemgående konceptuelt grundlag for kvalitetssikring.

Som eksempel på konsensus i monosurveysk litteratur om reliabilitet og validitet som basale kvalitetsbegreber kan Fowlers påpegning af, at pålidelige og gyldige målinger forudsætter ”gode” spørgsmål, citeres: ”A good question is one that produces answers that are reliable and valid measures of something we want to describe” (Fowler, 1995:2). Der er dog ikke fuldstændig konsensus om, hvordan begreberne defineres og heller ikke om til begreberne knyttede subbegreber (se fx Aldridge & Levine, 2001; de Vaus, 2002; Fink & Kosecoff, 1998; Fowler, 1993; Fowler & Mangione, 1990; Hellevik, 1993a; Sapsford, 1999). Mere påfaldende er det, at mange metodeforfattere tager de to begreber til efterretning uden at definere dem. Store dele af monosurveysk metodelitteratur er uden refleksioner om begreberne.

Men hvad er reliabilitet og validitet, og hvordan afgøres det, om et måleproblem er et reliabilitets- eller validitetsproblem? I mange fremstillinger opfattes reliabilitet som fravær af ”tilfældige målefejl” (se fx Hansen & Andersen, 2000; Hellevik, 1993a). Andre inddrager tidsdimensionen og påpeger, at reliabilitet implicerer konsistens, dvs. opnåelse af samme måleresultat ved gentagne målinger (se fx Aldridge & Levine, 2001; Babbie, 1990). Dog fordres selvsagt, at der ikke i mellemtiden er sket ændringer i de ”objekter”, som givne surveysspørgsmål angår (se også de Vaus, 2002).

Trods nuanceforskelle er den gennemgående opfattelse, at reliabilitet forudsætter, at måleinstrumenter genererer konsistente målinger. Det gælder fx også Fowler, for hvem reliabilitet er: ”... the extent to

which people in comparable situations will answer questions in similar ways” (Fowler, 1993:80). Der findes dog andre opfattelser. For Litwin (1995) er reliabilitet fx enbetydende med reproducerbarhed, mens andre definerer reliabilitet som fravær af ”målefejl”: ”A reliable survey instrument is one that is relatively free from ‘measurement error’ ” (Fink, 1995:46).

Hvad det andet kvalitetsbegreb – validitet – angår, betegnes det ofte som fravær af ”systematiske fejl” og handler om, hvorvidt spørgeskemaer og surveydata måler det tilsigtede: ”Validity refers to the degree to which a survey instrument assesses what it purports to measure” (Fink, 1995:49; se fx også Aldridge & Levine, 2001; de Vaus, 2002). Der findes imidlertid andre opfattelser, hvoraf nogle sætter fokus på det velkendte, men kontroversielle korrespondensprincip, dvs. forskningsresultaters overensstemmelse med ”virkeligheden”. Eksempelvis bifalder Fowler (1993), at en høj grad af validitet forudsætter, at det tilsigtede måles, men føjer til, at målinger skal korrespondere med *true values*, hvilket indebærer, at subjektive spørgsmål ikke lader sig validere. Som de fleste efterlyser også Sapsford måling af det tilsigtede, men uden reference til *true values*. Her er validitet “... the extent to which the data constitute accurate measurements of what is supposed to be measured” (Sapsford, 1999:9). Om end både validitet og reliabilitet virker som bærende kvalitetsbegreber i monosurveysk litteratur, er der således kun tendentiell konsensus om, hvad begreberne betyder.

Der er heller ikke enighed om med validitetsbegrebet forbundne subbegreber. I en dansk lærebog nævnes fire begreber: indholdsvaliditet (*content validity*), begrebsvaliditet (*construct validity*), kriterievaliditet (*criterion-related validity*) og prædiktiv validitet (Hansen & Andersen, 2000), mens den øvrige monosurveyske litteratur peger på yderligere subbegreber, ligesom de fire begreber ikke defineres på helt samme måde andetsteds (se fx Aldridge & Levine, 2001; de Vaus, 2002; Fink, 1995; Fowler, 1993; Hellevik, 1993a; Litwin, 1995; Olsen, 1998; Sapsford, 1999).

Ifølge Hansen & Andersen (2000) angår indholdsvaliditet overensstemmelse mellem teoretiske og operationelle definitioner af begreber, hvilket fx bifaldes af Babbie, ifølge hvem indholdsvaliditet “... refers to the degree to which a measure covers the range of meanings included in the concept” (Babbie, 1990:134). Andre peger på, at indholdsvaliditet: “... is a subjective measure of how appropriate the items seem to a set of

reviewers who have some knowledge of the subject matter” (Litwin, 1995:35).

Hvad begrebsvaliditet angår, sætter validitetsformen fokus på overensstemmelse mellem målingers faktiske og teoretisk forventede udfald (se fx Hansen & Andersen, 2000). En beslægtet opfattelse gøres gældende af Babbie, der pointerer, at begrebsvaliditet ”... is based on the way a measure relates to other variables within a system of theoretical relationships” (Babbie, 1990:134; se også Litwin, 1995). Aldridge og Levine (2001) mener derimod, at validitetsformen “... looks back at the performance of a measure over time ... to see if it has produced fruitful findings” (p. 40).

Kriterievaliditet angår overensstemmelse mellem målinger og foreliggende kriterievariabler, dvs. variabler, hvis anvendelighed forudsættes, fx fordi de er velkendte og afprøvede. Her gælder ligeledes, at der ikke er konsensus om, hvad validitetsformen implicerer (se fx Babbie, 1990; Fink, 1995; Judd et al., 1991; Litwin, 1995). Ét aspekt af kriterievaliditet er *concurrent validity* (”parallelvaliditet”), der sætter fokus på målingers grad af overensstemmelse med andre tilsvarende målinger af høj kvalitet. Hermed afprøves ”... how well the item or scale correlates with “gold standard” measures of the same variable” (Litwin, 1995:45).

Hvad derimod prædiktiv validitet angår, defineres begrebet stort set på samme måde i monosurveysk litteratur, dvs. som sammenligninger af forventede og faktiske fremtidige tilstande. Ifølge Litwin, fx, måler prædiktiv validitet “... how well the item or scale predicts expected future observations” (Litwin, 1995:45). Til de nævnte validitetsformer kan føjes umiddelbar validitet (*face validity*), hvorom Babbie bemærker: “Particular empirical measures may or may not jibe with our common agreements and our individual mental images ...” (Babbie, 1990:133), mens Fink mener, at begrebet angår, ”... how a measure appears on the surface” (Fink, 1995:49). Selv hermed er listen af begreber ikke tilendebragt, idet fx også nomologisk validitet og statistisk validitet anvendes i litteraturen. Hertil kommer endelig, at mange monosurveyske metodeforfattere undlader at pege på specifikke validitetsformer, men anvender validitetsbegrebet i dets almene form.

Sammenfattende gives der ikke konsensuelle svar på, hvad validitet og reliabilitet er, og hvorledes gyldighed og pålidelighed lader sig afprøve, ligesom mange forskere undlader at diskutere problematikken (se *oversigt 2.1.*). Der er derimod udbredt enighed om de to begreber som



konceptuelt grundlag for forebyggelse af måleproblemer ved spørgeskemakonstruktion og datatilvirkning. Ligeledes understreges ofte, at validitet og reliabilitet ikke er symmetriske (se fx Aldridge & Levine, 2001; Andersson, 1994; Fink, 1995; Hellevik, 2002; Sapsford, 1999). Ifølge en monosurveysk ”standardlærebog” kan målinger have en høj grad reliabilitet uden at være valide, mens det omvendte er umuligt. Som bl.a. påpeget af Hansen og Andersen (2000) kan det i praksis – fx fordi forskeres spørgeintention er tvetydig – være vanskeligt at afgøre, om et måleproblem er et validitets- eller reliabilitetsproblem. Derfor omtaler nogle forskere de to begreber under fællesbetegnelsen måleproblem (se fx Olsen, 2001a) eller *measurement error*.<sup>12</sup>

## Oversigt 2.1.

Opfattelser af reliabilitet og validitet i monosurveysk litteratur.

Reliabilitet	Validitet	Validitetsformer
Fravær af tilfældige fejl	Fravær af systematiske fejl	Indholdsvaliditet
Konsistente målinger	Målinger af det tilsigtede	Begrebsvaliditet
Stabile målinger	Nøjagtige målinger af det	Kriterievaliditet
Reproducerbarhed	tilsigtede	Parallelvaliditet
Fravær af målefejl	Korrespondens med ”sande	Prædiktiv validitet
Ikke-definitioner	størrelser”	Umiddelbar validitet
	Ikke-definitioner	Andre former
		Ikke-specifikationer

## 2.3. MULTISURVEYSK ÆKVIVALENS

Selv om monosurveysk metodelitteratur langt fra tilbyder konsensuelle svar på, hvad validitet og reliabilitet er, er der som nævnt næsten fuldstændig enighed om de to begreber som konceptuelt grundlag for forebyggelse af måleproblemer. Hvorvidt det forholder sig på tilsvarende måde, når fokus ændres til multisurveysk metodelitteratur, er problematikken i det følgende.

12. I en dansk afhandling defineres måleproblem som følger: Når en målings troværdighed anfægtes på grund af mangelfuld konstruktionsvaliditet, manglende muligheder for test af ekstern validitet, sandsynligt fravær af ekstern validitet og/eller på grund af empirisk sandsynliggjorte reliabilitetsproblemer, foreligger der et måleproblem (Olsen, 2001a).

En gennemgang af næsten 200 multisurveyske videnskabelige bøger og artikler viser, at begreberne reliabilitet og validitet er sjældent forekommende, mens ækvivalens anvendes som konceptuelt grundlag for forebyggelse af måleproblemer i langt de fleste tilfælde: "Much of the literature on cross-national research refers to the quest for 'equivalence' and how elusive it usually turns out to be" (O'Shea, R. et al., 2005:7). Eftersom multisurveys overskrider grænser og/eller kulturer, er fokusering på ækvivalens nærliggende, men forklarer ikke, hvorfor der er behov for et kvalitetsbegreb, som overskrider reliabilitet og validitet (se kapitel 6).

I de forholdsvis få tilfælde, hvor der henvises til de fra monosurveysk litteratur velkendte begreber reliabilitet og validitet, er det som regel i kombination med ækvivalensbegrebet, dvs. at reliabilitet og validitet ikke i sig selv er garant for sammenlignelige data, men fordrer udvidet fokusering på ækvivalens. Det gælder fx Behling og Law (2000), der påpeger, at også multisurveys skal tilfredsstille krav om validitet og reliabilitet, dvs. fravær af "tilfældige fejl" og "systematiske fejl", men desuden semantisk, konceptuel og normativ ækvivalens (se nedenfor). Det begrundes dog ikke, hvorfor supplerende kvalitetsbegreber er påkrævet. Der er sandsynligvis tale om særligt udviklede konceptuelle konventioner.

Hertil kommer yderligere et begreb, nemlig *bias*, der dog kun forekommer i enkelte bøger og artikler (se fx Van de Vijver, 1998; Van de Vijver & Leung, 1997). Bias, der fx kan være begrebslig og/eller metodologisk,<sup>13</sup> truer multisurveydatas sammenlignelighed og fører til manglende ækvivalens: "Bias refers to the presense of nuisance factors that challenge the comparability of scores across culturel groups ..." (Van de Vijver, i: Harkness et al., 2003:144). Van de Vijver påpeger, at bias og ækvivalens ikke er hinandens spejlinger, men forbundne kar: "The measurement implications of bias for comparability are adressed in the concept of equivalence" (Van de Vijver, i: Harkness et al., 2003:144). Ved reduktion af bias fremmes dermed ækvivalente data.

---

13. Ifølge Van de Vijver foreligger begrebslig bias (*construct bias*), når begreber ikke er identiske på tværs af kulturelle grupper, mens metodebias (*method bias*) fx angår spørgeskemaforskelle og interviewereffekter. Også *item bias* nævnes, dvs. når spørgsmål ikke virker på samme måde blandt respondenter (Van de Vijver, i: Harkness et al., 2003).

Trods enkeltes fremhævelse af biasbegrebet er begrebet ækvivalens dominerende i multisurveysk litteratur og opfattes af mange som afgørende: "Equivalence is the only meaningful criterion if data is to be compared from one context to another" (Hoffmeyer-Zlotnik & Wolf, i: Hoffmeyer-Zlotnik & Wolf, 2005:389). Der er dog ikke konsensus om begrebet og dets vekslende former. I en semantisk ordbog er ækvivalens "værende af samme værdi eller gyldighed", dvs. at surveydata er sammenlignelige over grænser og på tværs af kulturer (se fx van Herk et al., 2005; Patel et al., 2005). I multisurveysk metodelitteratur findes imidlertid ingen eller kun vage almene ækvivalensdefinitioner, fx: "For data to be equivalent it needs to have the same meaning, interpretation and accuracy in all countries and cultures" (van Herk et al., 2005). Derimod findes en lang række specifikke ækvivalensbegreber – ofte med vage definitioner, undertiden helt uden. Som fremstillingen nedenfor anskueliggør, vedrører begreberne ofte specifikke faser i den samlede forskningsproces (se fx Patel et al., 2005).

Nedennævnte eksempler er nogle af de oftest forekommende blandt over et halvt hundrede former for ækvivalens (se fx Blair & Piccinino, i: Hoffmeyer-Zlotnik & Harkness, 2005; Johnson, 1998) (se *oversigt 2.2*). Nogle af begreberne grupperes undertiden i overordnede ækvivalens kategorier, men uden almene ækvivalensdefinitioner. Eksempelvis grupperer Johnson (1998) en mangefold af ækvivalensformer i to overordnede kategorier, nemlig fortolkende versus procedural ækvivalens. Herom senere, men først en præsentation af ofte forekommende ækvivalensbegreber.<sup>14</sup>

---

14. Hertil kommer yderligere – sjældnere i multisurveysk litteratur nævnte – ækvivalensbegreber, fx dataindsamlingsækvivalens, undertiden kaldet procedural ækvivalens (*technical equivalence*), indikatorækvivalens (*indicator equivalence*), indholdsækvivalens (*content equivalence*), kontekstuel ækvivalens (*contextual equivalence*), kriteriel ækvivalens (*criterion equivalence*) og stikprøveækvivalens (*sample equivalence*) (se fx Herdman et al., 1997; Hui & Triandis, 1989; Johnson, 1998, og Singh et al., 2000).

## Oversigt 2.2.

Opfattelser af ækvivalens i multisurveysk litteratur.

Funktionel ækvivalens ( <i>functional equivalence</i> )	Sammenlignelige fænomener
Kalibreringsækvivalens ( <i>calibration equivalence</i> )	Sammenligneligheds- skabende måleenhedsjusteringer
Kategoriel ækvivalens ( <i>categorical equivalence</i> )	Sammenlignelige objekt- grupper, fx produktkategorier.
Konceptuel ækvivalens ( <i>conceptual equivalence</i> )	Sammenlignelige begreber
Metrisk ækvivalens ( <i>metric equivalence</i> )	Sammenlignelige skala-/ scoreprocedurer
Måleenhedsækvivalens ( <i>measurement unit equivalence</i> )	Sammenlignelige måleenheder
Normativ ækvivalens ( <i>normative equivalence</i> )	Sammenlignelighed mht. samfundsmæssige regler
Oversættelsesækvivalens ( <i>translation equivalence</i> )	Samme tværgående betydninger
Responsækvivalens ( <i>response equivalence</i> )	Ikke defineret (vedr. respon- denteres opgaveløsninger)
Semantisk ækvivalens ( <i>semantic equivalence</i> )	Som oversættelsesækvivalens
Skalaækvivalens ( <i>scalar equivalence</i> )	Sammenlignelige skalaer/ ratio scale-sammenlignelighed
Strukturel ækvivalens ( <i>structural equivalence</i> )	Sammenlignelige begreber/ samme faktorstrukturer

Begrebet funktionel ækvivalens – et af de oftest nævnte begreber – omhandler, hvorvidt fænomener, som problemstillinger sætter fokus på, spiller samme rolle i forskellige lande og/eller kulturer, og opfattes af mange som en basal forudsætning for tværkulturelle sammenligninger: “... functional equivalences must pre-exist as naturally occurring phenomena; they are discovered and cannot be created or manipulated by the cross-cultural researchers. Without this equivalence ... no valid cross-cultural behavioral comparisons may be made” (Berry, 1969:122). I

så henseende bliver et afgørende spørgsmål, om multisurveys måler samme sociale eller andre fænomener på tværs af grænser og/eller kulturer (se fx Hoffmeyer-Zlotnik & Wolf, i: Hoffmeyer-Zlotnik & Wolf, 2005). Funktionel ækvivalens opnås, "... when two or more behaviors (in two or more cultural systems) are related to functionally similar problem" (Tran & Williams, 1994:3).

Ifølge van Herk et al. (2005) retter funktionel ækvivalens sig imod, hvorvidt "... the concepts, objects or behaviours studied have the same role or function in all countries included in the analysis" (p. 5). Som eksempel kan nævnes, om cykler udelukkende eller overvejende virker som transportmiddel eller som rekreative hjælpemidler: "In China, for example, we're dealing with an everyday mode of transport, while in the USA, the bicycle is more of a fitness and sporting tool, which means that the competitive set may be different in these two countries" (Patel et al., 2005:8).

Yderligere en opfattelse af funktionel ækvivalens betoner teoretiske dimensioner. Funktionel ækvivalens opnås i så fald, "... when questions asked in the different countries stand in identical relationships to the intended theoretical dimensions" (Alwin et al., 1994:30). Selv om funktionel ækvivalens basalt handler om, hvorvidt fænomener spiller samme rolle på tværs af lande og kulturer, er der adskillige opfattelser af, hvad funktionel ækvivalens er (eksempler, se fx Berry, 1969; Hui & Triandis, 1989; Johnson, 1998; Pareek & Rao, 1980; Singh et al., 2000; Teune, 1990), ligesom flere multisurveyske forskere påpeger vanskelighederne ved opnåelse af funktionel ækvivalens. Harpaz (2000) nævner følgende eksempel: "As a case in point, if one wishes to compare the concept of work centrality in the US, Japan, and Israel, ascertaining its functional equivalence requires assessment of construct functions in all three countries. Performing such an assessment necessitates a thorough conceptualization and operationalization of work centrality in these nations. To realize functional equivalence as a construct, work centrality must mean and encompass equivalent functions in the US, Japan, and Israel, as substantiated by analogous nomological networks of relationships both between and within constructs".

De to følgende begreber – kalibreringsækvivalens og kategoriel ækvivalens – nævnes markant sjældnere i multisurveysk litteratur end funktionel, men beskrives, fordi begreberne inkluderes i mere overordnede ækvivalens kategorier (se nedenfor). Kalibreringsækvivalens vil sige

at justere eller tilpasse måleenheder, hvorved de gøres sammenlignelige over grænser eller kulturer, fx valuatajusteringer, farver med samme betydning etc. (se fx Patel et al., 2005; van Herk et al., 2005). Kategoriel ækvivalens handler om, hvorvidt grupper af objekter – fx produktkategorier – er sammenlignelige over landegrænser eller på tværs af kulturer (se fx van Herk et al., 2005). Det kan fx være, om indtagelse af øl både virker som *soft drink* i Syd- og Nordeuropa, hvilket ikke synes at være tilfældet (se fx Patel et al., 2005).

Kalibreringsækvivalens er beslægtet med måleenhedsækvivalens, der forekommer i dele af multisurveysk litteratur (se fx Berry et al., 1997; Douglas & Nijssen, 2002; Johnson, 1998; Patel et al., 2005). Denne ækvivalensform handler om to eller flere måleenheders sammenlignelighed: "Instruments (dvs. spørgeskemaer, h.o.) show this if their measurement scales have the same units of measurement, but a different origin" (Van de Vijver, i: Harkness et al., 2003:153). Et eksempel er forskellige temperaturskalaer, der uden videre lader sig "oversætte", fx Fahrenheit- og Celcius-skalaen. Også omregning af valuater kan nævnes, hvorved det konceptuelle overlap mellem kalibrerings- og måleenhedsækvivalens er anskueliggjort.

Adskillige forskere anskuer konceptuel ækvivalens – et af de oftest nævnte begreber – som grundlæggende forudsætning for sammenligninger over grænser såvel som mellem kulturer (se fx Behling & Law, 2000; Harpaz, 2000; Johnson, 1998; Okazaki & Sue, 1995; Patel et al., 2005; Van de Vijver, i: Harkness et al., 2003; van Herk et al., 2005; Willis, i: Cohen & Lepkowski, 2004). Konceptuel ækvivalens vil fx sige, at et begreb indeholdt i en problemstilling har samme betydning over landegrænser og/eller på tværs af kulturer: "... a construct may be considered conceptually equivalent in two or more cultures when it can be meaningfully discussed in each culture and it has a comparable meaning across all cultures involved" (Harpaz, 2000). Eller med en tilsvarende formulering af van Herk et al. (2005): "Conceptual equivalence requires the interpretations that individuals place on objects, stimuli or behaviour to be consistent across cultures" (p. 353). En beslægtet opfattelse gøres gældende af Tran og Williams (1994), ifølge hvilke konceptuel ækvivalens vedrører, "... the comparability of the meaning of research materials across research groups" (p. 3). Ækvivalente begreber er med andre ord begreber, der er fælles for flere lande og/eller kulturer.

Andre påpeger, at konceptuel ækvivalens indebærer fælles tværkulturelle referencerammer: "... conceptual equivalence refers to the degree to which members of different cultures use a common frame of reference when responding to items on a survey instrument" (Schaffer & Riordan, 2003:191). Ifølge Hui & Triandis (1989) og andre indebærer konceptuel sammenlignelighed ikke, at et givet begreb har eksakt samme betydning på tværs af grænser og kulturer, da begrebet kan have forskellige aspekter og dimensioner. Det skal blot være muligt, at begrebet kan blive *meaningfully discussed*. I lighed med andre begreber er der således ikke konsensus om, hvordan konceptuel ækvivalens defineres: "Certain authors have suggested it has to do with existence of the concept in another culture, whilst others have suggested that it has to do with whether items in the source and target versions of a scale are similar in meaning" (Herdman et al., 1997:243).

Et sjældnere begreb beslægtet med konceptuel ækvivalens er strukturel ækvivalens (se fx Harkness et al., 2003; Johnson, 1998). Trods sjældent nævnt i litteraturen omtales begrebet, fordi det inkluderes i en mere overordnet ækvivalenskategori. Også strukturel ækvivalens handler om, hvorvidt samme begreb måles på tværs af lande og kulturer: "An instrument administered in different cultural groups shows structural equivalence if it measures the same construct in all these groups" (Van de Vijver, i: Harkness et al., 2003:153). Van de Vijver og Leung drager konsekvensen og fordrer ensartede tværnationale og/eller kulturelle faktorstrukturer for at hævde strukturel ækvivalens. For dem henviser strukturel ækvivalens til "... the similarity of psychometric properties of data from different cultures" (Van de Vijver & Leung, i: Berry et al., 1997:261).

Begrebet metrisk ækvivalens, der er beslægtet med skalaækvivalens, og som ofte omtales i forbindelse hermed, retter sig ifølge van Herk et al. (2005) imod "... specific scale or scoring procedure used for assessment" (p. 353). Men heller ikke herom foreligger der konceptuel konsensus. Eksempelvis påpeger Harpaz (2000): "Metric equivalence prevails when the psychometric values in different populations representing several data sets (cultures or nations) demonstrate virtually similar composition or adherence"<sup>15</sup> (se også Tran & Williams, 1994), mens

---

15. Harpaz (2000) skriver videre: "Basically, there are two requirements for establishing metric equivalence (...). The first demands that statistical association among dependent and independ-

andre ikke her nævnte definitioner af metrisk ækvivalens afviger fra ovennævnte.

Med fokus på skalaækvivalens rettes opmærksomheden imod, om kulturelt forskellige respondenter opfatter skalaer på samme måde, fx Likert-skalaer: "... an essential ingredient in any comparison is a variable that forms a scale which is identical or invariant scale properties for the persons or groups to be compared" (Poortinga, 1989:737) Ifølge Schaffer & Riordan (2003) handler skalaækvivalens derfor ligeledes om, "... whether respondents perceive and interpret rating-scale intervals in the same manner" (p. 191). Schaeffer & Riordan anskueliggør eksempelvis, at på midten-svaret 3 i en 5-punkts Likert-skala både kan betyde *no opinion* og *mild agreement*. En tilsvarende opfattelse af skalaækvivalens gøres fx gældende af Douglas & Nijssen (2002): "Scalar equivalence requires that response to a given scale is equivalent or has the same meaning and interpretation across contexts".

Fordringer til skalaækvivalens kan imidlertid være mere vidtgående og rette sig imod datasammenlignelighed på måleniveauer, der fx muliggør beregning af gennemsnit, t-test og variananalyse: "Scalar equivalence assumes the identical interval or ratio scales across cultural groups" (Van de Vijver, i: Harkness et al., 2003:154). Van de Vijver og andre betragter tilmed skalaækvivalens som den "højeste" form for ækvivalens og betegner undertiden skalaækvivalens som *full score comparability* (se fx også Van de Vijver & Leung, 2000; Van de Vijver & Leung, i: Berry et al., 1997).

Yderligere et begreb, der dog kun forekommer sporadisk, men som nævnes her, fordi det inkluderes i en mere overordnet kategori, er normativ ækvivalens, der fx defineres som: "... the degree to which the researcher has dealt successfully with the problems created by differences in societal rules ..." (Behling & Law, 2000:16). Den slags regler handler fx om åbenhed eller mangel på samme, hvormed bestemte emner disku-

---

ent variables stay relatively stable, regardless whether the obtained variance is used within a state/culture or internationally. According to this approach, co-variation among variables should be constant irrespective of the cause of variation. The second requirement is that before any comparisons are made, statistical relationships among dependent variables should be formed congruently in the involved nations/cultures. This may be evinced via similarity in correlation matrices or by common factor structures (...). Measurement instruments should be designed in a similar manner within cultures/nations before any international comparisons are drawn. Ordinarily, it is possible to institute metric equivalence only after completion of data collection and analysis".



teres, fx politiske, og måder hvorpå ideer og forestillinger udtrykkes, fx direkte eller indirekte.

Multisurveys indebærer normalt, at spørgeskemaer foreligger på to eller flere sprog. Det er derfor ikke overraskende, at også oversættelsesækvivalens, der er tæt knyttet til konceptuel ækvivalens, nævnes i multisurveysk litteratur. I multisurveys forudsætter uanfægtelige målinger, at spørgsmål på ét sprog kan oversættes, så de bevarer deres betydning på andre sprog (se fx Harkness, 1999; Iyengar, 1983). Oversættelsesækvivalens vil derfor fx sige, at spørgsmål ikke blot er forståelige, men også har samme betydning i forskellige nationale og/eller kulturelle sammenhænge (van Herk et al., 2005). Oversættelsesækvivalens har forskellige aspekter, fx leksikalske og grammatiske: "Translation equivalence can be divided into lexical, idiomatic, grammatical-syntactical and experiential equivalencies" (Usunier, 1999). Mens det leksikalske aspekt angår, om ord uden videre er oversættelige, angår det grammatiske fx ordsammensætninger.

I dele af litteraturen betegnes oversættelsesækvivalens semantisk ækvivalens (se fx Behling & Law, 2000; Herdman et al., 1997; Johnson, 1998). Enkelte anvender dog benævnelsen lingvistisk ækvivalens (se fx Iyengar, 1983). Eksempelvis forudsætter semantisk ækvivalens for Behling og Law (2000), at "... the meaning of the source language statement is preserved in the translation" (p. 15). Tilsvarende opfattelse findes andetsteds, fx hos Singh et al. (2000), der påpeger, at semantisk ækvivalens indebærer, at spørgsmål har uændret betydning efter oversættelse og anvendelse.

Blandt de nævnte ækvivalensformer resterer én, nemlig responsækvivalens, der ikke defineres i den litteratur, som rapporten baseres på. Responsækvivalens handler om respondenters kulturelt betingede og derfor undertiden markant forskellige opgaveløsninger, der påvirker surveydatas sammenlignelighed (se fx Baumgartner & Steenkamp, 2001; Smith, 2003). Det drejer sig fx om respondenters tendenser til på midten- versus pol-svar og "socialt ønsket" svarredigering (se kapitel 4).

Nogle af de tolv præsenterede ækvivalensbegreber grupperes undertiden i overordnede ækvivalens kategorier, men uden almene ækvivalensdefinitioner. Af fremstillingen fremgår, at der ikke er konsensus om, hvordan ækvivalens konceptualiseres, hvilket også gælder gruppering af ækvivalensbegreber i overordnede kategorier. For de fleste bøger og

artikler, som rapporten baseres på, gælder, at sådanne grupperinger undlades, hvorfor kun få eksempler kan nævnes (se *oversigt 2.3.*).

### Oversigt 2.3.

Overordnede ækvivalens kategorier.<sup>1)</sup>

	Overordnede kategorier	Kategoriens bestanddele
1. gruppering	Fortolkende ækvivalens ( <i>interpretive equivalence</i> )	Konceptuel ækvivalens Funktionel ækvivalens Semantisk ækvivalens (flere eksempler nævnes)
	Procedural ækvivalens ( <i>procedural equivalence</i> )	Metrisk ækvivalens Måleenhedsækvivalens Leksikalsk ækvivalens <sup>2)</sup> (flere eksempler nævnes)
2. gruppering	Konstruktionsækvivalens ( <i>construct equivalence</i> )	Funktionel ækvivalens Konceptuel ækvivalens Kategoriell ækvivalens
	Måleækvivalens <sup>3)</sup> ( <i>measurement equivalence</i> )	Kalibreringsækvivalens Oversættelsesækvivalens Metrisk ækvivalens
3. gruppering	Generel ækvivalens (ikke defineret)	Konstruktionsækvivalens (se ovenfor) Måleenhedsækvivalens Skalaækvivalens
4. gruppering	Generel ækvivalens (ikke defineret)	Semantisk ækvivalens Konceptuel ækvivalens Normativ ækvivalens

1) 1. gruppering, se Johnson, 1998; 2. gruppering, se van Herk et al., 2005; Patel et al., 2005; 3. gruppering, se Van de Vijver & Leung; 1997; samme, 2000; 4. gruppering, se Behling & Law, 2000.

2) Ækvivalensformen er ikke nævnt tidligere.

3) Hertil kommer dataindsamlingsækvivalens, der er ekskluderet af oversigten, fordi rapporten ikke sætter fokus på dataindsamlingsformer.

I en udmærket oversigt over mere end 50 ækvivalensbegreber peger Johnson (1998) på en grundlæggende, men diffus sondring mellem fortolkende ækvivalens og procedural ækvivalens. Det er en sondring, som bifaldes i dele af multisurveysk litteratur (se fx Hudler & Richter, 2000). Efter at have redegjort for de fleste af begreberne og påpeget deres upræcise eller fraværende definitioner mener Johnson at kunne gruppere dem i de to nævnte hovedkategorier. Fortolkningsækvivalens handler især om ords og begrebers betydning, hvorfor konceptuel ækvivalens er

vigtig, dvs. at begreber er relevante og meningsfulde i lande eller inden for kulturer, som en given undersøgelse sætter fokus på. Derudover forudsætter betydningsoverensstemmelse fx også funktionel og semantisk ækvivalens. På den anden side angår procedural ækvivalens: "... measures and procedures used to make cross-national comparisons" (Johnson, 1998:7), fx metrisk ækvivalens, måleenhedsækvivalens og leksikalsk ækvivalens. Også van Herk et al. (2005) m.fl. sonderer mellem to overordnede ækvivalens-kategorier, men differentierer ikke som Johnson (1998) (se *oversigt 2.3.*). For dem er sondringen mellem konstruktionsækvivalens og måleækvivalens. Førstnævnte handler om, "... whether the construct in question has the same significance and importance across different cultures" (op.cit.), hvilket udelukkende er muligt, såfremt målinger både opfylder krav om konceptuel, kategoriel og funktionel ækvivalens. Måleækvivalens handler derimod om, hvorvidt måleinstrumenter – dvs. spørgeskemaer – genererer nøjagtige og sammenlignelige surveydata på tværs af lande og/eller kulturer, hvilket fordrer opfyldelse af andre tre betingelser: oversættelsesækvivalens, kalibreringsækvivalens og metrisk ækvivalens.

De to sidste grupperinger er kendetegnet ved, at tre specifikke ækvivalensformer er afgørende for, at der foreligger generel – men i øvrigt ikke nærmere defineret – ækvivalens. Hos Van der Vijver og Leung (1997; 2000) er de tre ækvivalensformer placeret i et hierarki med konstruktionsækvivalens (*construct equivalence*)<sup>16</sup> nederst, dvs. som grundlæggende forudsætning for sammenlignelighed. Et højere ækvivalensniveau er måleenhedsækvivalens, mens det højeste er skalaækvivalens: "This level of equivalence can be achieved when the measurement instrument is on the same ratio scale in each cultural group" (Van de Vijver & Leung, 1997:8).

Hvad den sidste gruppering angår, er den særligt udviklet i tilknytning til oversættelse af spørgeskemaer (Behling & Law, 2000) – og måske derfor mindre relevant end de foregående. Afsluttende skal derfor blot nævnes, at generel sammenlignelighed ifølge Behling og Law både forudsætter semantisk, konceptuel og normativ ækvivalens, hvor den sidste form som nævnt handler om, hvorvidt der ved gennemførelse af

---

16. Konstruktionsækvivalens omfatter sandsynligvis både konceptuel, kategoriel og funktionel ækvivalens, hvilket dog ikke ekspliciteres. Ifølge Van de Vijver & Leung (1997) fordrer konstruktionsækvivalens, at samme *construct* måles tværkulturelt.

multisurveys tages højde for samfundsmæssige reglers forskelligheder  
(*societal rules*).

# SPØRGESKEMA- KONSTRUKTION

## 3.1. INDLEDNING

Rapportens anden problemstilling sætter fokus på, hvad multisurveysk litteratur bidrager med vedrørende formulering og oversættelse af spørgsmål:

- Hvordan formuleres anvendelige spørgsmål over grænser eller på tværs af kulturer, og hvordan oversættes spørgsmål i multisurveys til forebyggelse af måleproblemer?

Også ved belysningen af denne problemstilling sættes der fokus på forskelle mellem multi- og monosurveysk litteratur. Kapitlet indledes derfor med en summarisk redegørelse for fremherskende anbefalinger vedrørende spørgeskemakonstruktion og spørgsmålsformulering i monosurveys (afsnit 3.2.). Det drejer sig næsten udelukkende om erfaringsbaserede *common sense*-anbefalinger, idet forskningsbaserede anbefalinger, der tager afsæt i respondenteres semantiske og andre kognitive opgaveløsninger, henvises til det efterfølgende kapitel.<sup>17</sup> Derpå belyses, hvad multisur-

---

17. Ved forskningsbaserede anbefalinger forstås anbefalinger, der baseres på empiriske undersøgelser, fx kognitive interview og *split sample*-eksperimenter med respondenter.

veysk litteratur bidrager med i så henseende (afsnit 3.3.). I sagens natur handler multisurveys i langt højere grad end monosurveys om oversættelse af spørgeskemaer, der behandles i et særskilt afsnit (afsnit 3.4.).<sup>18</sup> Det drejer sig eksempelvis om oversættelsesmuligheder og -procedurer. Forskelle og ligheder mellem mono- og multisurveysk litteratur sammenfattes og problematiseres i rapportens afsluttende kapitel (afsnit 6.2.)

### 3.2. MONOSURVEYSKE ANBEFALINGER

Er konstruktion af spørgeskemaer med Paynes (1951) legendariske formulering *the art of asking questions* eller *the interdisciplinary science of asking questions?* (Olsen, 2001a, bd. II:200). Er spørgeskemakonstruktion common sense eller kan den "videnskabeliggøres"? Hvordan muliggøres formulering af "gode" survey spørgsmål og "gode" spørgeskemaer? Det er problematikken i det følgende, hvor der så vidt muligt sættes fokus på spørgeskemakonstruktion som sådan, mens respondenteres svarprocessuelle opgaveløsninger inddrages i kapitel 3. Derfor omhandler afsnittet især monosurveyske "tommelfingerregler".

Efter mange monosurveyske forfatters opfattelse er spørgeskemakonstruktion surveys "kerne" og afgørende for måleproblemer: "The questions asked of respondents are the ultimate core of the survey project" (Alreck & Settle, 1995:87). Der findes en righoldig litteratur med råd om, hvordan "gode" spørgsmål og spørgeskemaer udarbejdes (se fx Aldridge & Levine, 2001; Alreck & Settle, 1995; Babbie, 1990; Bickman & Rog, 1998; Fink, 1995; Folz, 1996; Fowler, 1993/1995; Frey & Oishi, 1995; Judd et al., 1991; Kumar, 1999; Neuman, 2000; Peterson, 2000; Rea & Parker, 1997). De aspekter, som anbefalingerne retter sig imod, sammenfattes i *oversigt 3.1.*

Efter nogle metodeforfatteres opfattelse er spørgeskemakonstruktion ikke blot surveyundersøgelsers "kerne", men en vanskelig opgave: "We need to frame questions that are meaningful, sensitive, precise, searching, and salient to our respondents" (Aldridge & Levine, 2001:95). Hvordan gøres det i praksis? Selv om der synes at være overvejende konsensus om mange kriterier for spørgeskemakonstruktion, er

---

18. Oversættelse af spørgeskemaer kan også komme på tale i monosurveys, fx hvis nogle respondenter ikke behersker populationens fremherskende sprog.

der, når monosurveysk litteratur anskues under ét, ikke enighed om, hvordan ”gode” spørgeskemaer udarbejdes, ligesom mange kriterier er diffuse og kun lidt operationelle.

### Oversigt 3.1.

Aspekter vedrørende spørgeskemakonstruktion ifølge dele af monosurveysk metode-litteratur.

Spørgsmål	Svarkategorier	Spørgeskema
Temarelatering	Tilvalg/fravalg	Spørgsmålsrækkefølge
Definitioner	Spørgsmålsrelatering	Omfang
Operationalisering	Antal	Formattering
Grammatik	Balancering	Andre aspekter
Fokusering	Ved ikke-kategorier	
Længde	På midten-kategorier	
Formulering	Andre aspekter	
Neutralitet		
Relevans		
Hukommelse		
Tidsrum		
Andre aspekter		

Om formulering af spørgsmål peger en del forfattere på, at spørgsmål i et spørgeskema skal relateres eksplicit til problemstillinger (se fx Fowler, 1995). Eller med de Vaus’ ord: ”The difficulty is in developing good indicators for concepts” (de Vaus, 1998:47; se også Czaja & Blair, 1996; Peterson, 2000), dvs. at operationalisere begreber: ”...most concepts of interest to social researchers have no real meaning, no ultimate definitions” (Babbie, 1990:119). Heraf følger ”dataindsamlingens” umulighed. Data indsamles aldrig, men konstrueres (se fx Hox, i: Lyberg et al., 1997).

Når begreber er defineret, er de ikke operationaliseret: “Operationalization is the process whereby researchers specify empirical observations that can be taken as indicators of the attributes contained within a given concept” (Babbie, 1990:121). Operationalisering handler om ”operationer”, der skal udføres for at tage stilling til, om et fænomen falder ind under eller tilhører et givet begreb (Hellevik, 1993a). Derfor peger flere forfattere på en særlig operationaliseringsfase (se fx de Vaus, 1998). Fasen kan være kritisk, fordi hverken problemstillinger eller begrebsdefinitioner umiddelbart er bestemmende for de spørgsmål, der skal formuleres og placeres i spørgeskemaet, og som genererer ”målelighed”.

Derfor skal forskere argumentere for, at spørgsmål er afledt af og relateret til problemstillinger og dertil hørende begreber. Efter hvilke retningslinjer begrebsindikatorer udvikles, giver monsoosurveysk litteratur dog hverken konsensuelle eller éntydige svar på: ”Det overordnede kriterium på en god operasjonalisering er begrepsvaliditet, defineret teoretisk og begrepet slik vi har lykket med å operasjonalisere det” (Kleven, i: Lund, 2002:183). Det er under operationaliseringsprocessen, at indholdsvaliditet tilsikres, hvilket er en kompliceret opgave.

Der advokeres desuden for, at spørgsmål formuleres grammatisk korrekt (se fx Alreck & Settle, 1995; Fink, 1995; Fowler, 1993). Anbefalingen synes evident, men er ikke nødvendigvis i overensstemmelse med en fordring om, at spørgsmåls afsæt er ”hverdagssprog”, der ikke altid kendetgnes ved korrekt grammatik. Ét grammatisk aspekt, som flere advarer imod, er anvendelse af negationer, ikke mindst dobbelte negationer (se fx Andersson, 1994; Converse & Presser, 1986; Oppenheim, 1992).

I dele af litteraturen peges endvidere på, at spørgsmål skal være éntydigt fokuserende (se fx Alreck & Settle, 1995; Babbie, 1990; de Vaus, 1998; Fink, 1995; Neuman, 2000; Oppenheim, 1992). Fokusering kræver, påpeger flere forfattere, at spørgsmål er éndimensionelle, dvs. kun retter sig imod ét emne: “Double-barreled questions are ambiguous and produce misleading and inaccurate answers” (Nolinske, 1998). Hvis et spørgsmål sætter fokus på flere emner, bør det omformuleres eller opdeles i to eller flere spørgsmål, før det kan tildeles en plads i spørgeskemaet.

Spørgsmålslængde er yderligere et aspekt. Selv om adskillige forskningsresultater problematiserer anbefalingen af spørgsmål med få ord (se kapitel 3), er efterlysningen af *brevity* en hyppig common sense-anbefaling i litteraturen (se fx de Vaus, 1998; Alreck & Settle, 1995; Babbie, 1990). Under overskriften *keep it simple* menes lange spørgsmål at være krævende og forvirrende. Nogle forfattere angiver ligefrem en øvre grænse for, hvor mange ord spørgsmål bør indeholde, fx maksimalt 20 eller 25 ord (se fx Oppenheim, 1992; Peterson, 2000; Rossi et al., 1983).

En fordring, som mange tildeler forrang, er sproglig klarhed og utvetydighed: “The meaning of the question must be completely clear to all respondents” (Alreck & Settle, 1995:89). Hvis spørgsmål ikke er éntydige, skabes måleproblemer. Det er derfor afgørende, at forskere formulerer spørgsmål, der ikke er flertydige: ”The best way to avoid ambiguity is to use short, crisp, simple questions” (de Vaus, 1998:84). Hvordan



éntydige spørgsmål formuleres, indeholder litteraturen dog kun få ledetråde for.

Ét vigtigt aspekt af spørgsmålsformulering er valg af ord, påpeger flere (se fx Alreck & Settle, 1995; Babbie, 1990; de Vaus, 1998; Frey & Oishi, 1995; Judd et al., 1991; Nolinske, 1998). Alreck & Settle (1995) anbefaler fx, at flertydige ord fravælges, hvilket indebærer, at der kun gøres brug af ”kerneord” (*core vocabulary*). Det er dog tvivlsomt, om brug af ”kerneord” tilsikrer pålidelige målinger. Som eksempler herpå nævner Singleton et al. (1988) ord som sjældent, mange og få.

Atter andre pointerer, at ord bør være uden ”bias”, fx uden negative konnotationer (se fx Judd et al., 1991). Da det imidlertid næppe altid er muligt at undgå flertydige eller ”svære” ord, anbefaler nogle forfattere, at ord eller termer defineres eller eksemplificeres (se fx Fowler, i: Bickman & Rog, 1998; Dahmström, 1996). Sidstnævnte kan selvsagt komme i konflikt med et forlangende om korte spørgsmål.

Yderligere en anbefaling er neutrale og balancerede spørgsmål, dvs. fravalg af ledende og bias-behæftede spørgsmål (se fx Alreck & Settle, 1995; Andersson, 1995; Babbie, 1990; de Vaus, 1998; Peterson, 2000). Her og andetsteds er problemet imidlertid, at afsender- og modtagerperspektivet ikke nødvendigvis er sammenfaldende. Hvad forskeren forstår som ikke-ledende, kan respondenter meget vel opfatte anderledes (se kapitel 4).

Nogle forfattere anbefaler tillige, at spørgsmål formuleres, så de muliggør respondenteres besvarelse, dvs. at de oplever spørgsmål som relevante og kan genkalde ønskede informationer (se fx Babbie 1990; Alreck & Settle, 1995; de Vaus, 1998). Derfor bør spørgsmål rette sig imod respondenteres egne erfaringer: ”The strength of survey research is asking people about their firsthand experiences” (Fowler, i: Bickman & Rog, 1998:365). Heraf følger, at fx hypotetiske spørgsmål og kausalspørgsmål bør fravælges (se fx Converse & Presser, 1986). Hermed øges sandsynligheden for, at respondenter kun ytrer sig om forhold, som de selv har erfaret.

Yderligere en måde, hvorpå pålidelig genkaldelse af informationer kan fremmes, er ved kun at spørge til nutidige fænomener eller til fænomener i den nærmeste fortid (se fx Fowler, 1995). Hvordan spørgsmål formuleres på måder, der – når der ses bort fra kundskabsspørgsmål – muliggør genkaldelse af ”sande” informationer, tilbyder

håndbøger og anden common sense-baseret metodelitteratur dog sjældent svar på.

Blandt supplerende anbefalinger kan nævnes, at der ofte er behov for, at spørgsmål præciseres i form af klare tidsafgrænsninger (se fx Fowler, 1995). Hertil kommer påpegning af, at spørgsmål ikke bør være "truende" eller nedladende (se fx Andersson, 1994), og at de heller ikke tildeles prestige bias (se fx de Vaus, 1998). Endelig anbefales, at spørgsmål afholder sig fra virkelighedsfjerne antagelser om respondenter, fx ved at forudsætte at alle respondenter er erhvervsaktive (se fx Judd et al., 1991; Neuman, 2000). Sidstnævnte problem løses almindeligvis ved anvendelse af filterspørgsmål.

Også hvad spørgsmåls "forlængelse" – svarkategorier – angår, indeholder litteraturen forskellige anbefalinger. Ét problem er, om spørgsmål skal efterfølges af svarkategorier eller ej. Mens nogle forfattere anbefaler lukkede spørgsmål, har andre mere nuancerede opfattelser. Eksempelvis anbefaler de Vaus (2002) og Folz (1996) lukkede spørgsmål, mens fx Converse & Presser reflekterer pro et contra: "Use close-ended and open-ended items for appropriate inquiries" (Converse & Presser, 1986). Trods forskellige opfattelser er surveyspørgsmål normalt lukkede. I så fald bør svarkategorier være sammenhængende med spørgsmålet, dvs. udgøre en "logisk" og betydningsmæssig forlængelse heraf (se fx Babbie, 1990; de Vaus, 2002; Fowler, 1995; Judd et al., 1991; Neuman, 2000). Nødvendige forudsætninger er da, at svarkategorier er udtømmende og gensidigt eksklusive. Om sidstnævnte kriterier synes der ikke overraskende at være konsensus, men ikke om, hvordan det tilsikres, at svarkategorier udgør "logiske" og betydningsmæssige spørgsmålsforlængelser.

Blandt andre problematikker er antallet af svarkategorier og balance disse imellem. Hvad den første angår, er anbefalingerne tvetydige. Megen monosurvey-litteratur er uden anbefalinger, mens andre peger på, at mængden af svarkategorier "afhænger af" omstændighederne, men bør være mindst mulig (se fx Converse & Presser, 1986). Atter andre pointerer, at antallet beror på, hvordan data tilvirkes, fx ved besøg eller telefon (se fx de Vaus, 2002). Med hensyn til den anden problematik gør adskillige opmærksom på, at svarkategorier bør balanceres, når dette er påkrævet, fx i holdningsspørgsmål med evaluerende kontinua. I modsat fald behæftes målinger med bias (se fx Andersson, 1994; de Vaus, 2002; Neuman, 2000).

Yderligere problematikker angår brug af ved ikke- og på midten-svarkategorier, hvoraf sidstnævnte især kommer på tale i tilknytning til holdningsspørgsmål med evaluerende kontinuua. Heller ikke her er anbefalingerne konsensuelle, men snarere ”det kommer an på”-pragmatiske. Mens nogle forfattere er af den opfattelse, at ved ikke-kategorier så vidt muligt bør undgås, mener andre, at den slags svar – navnlig i tilknytning til holdningsspørgsmål – er vigtige og anvendelige informationer (se fx Andersson, 1994; de Vaus, 2002; Johannessen & Tufte, 2002; Oppenheim, 1992).

Hverken spørgsmål eller svarkategorier udgør et spørgeskema. Først når spørgsmål udgør en helhed, har forskeren konstrueret et spørgeskema: "A questionnaire is organic, with each part vital to every other part, and all parts must be handled simultaneously to create this whole instrument" (Labaw, 1980:12). Dannelse af en spørgeskemahelhed rejser flere spørgsmål, der skal tages stilling til, fx: I hvilken rækkefølge skal spørgsmål placeres? Hvor mange spørgsmål skal skemaet inkludere? Hvordan skal spørgeskemaet formateres? Hvad rækkefølge angår, peger flere på, at den kan have afgørende betydning: "A poorly organized questionnaire can confuse respondents, bias their responses, and jeopardize the quality of the entire research effort" (Rea & Parker, 1997:35).

I så henseende indeholder monosurveysk litteratur forskellige i nogen grad konkurrerende råd. Én gennemgående opfattelse synes at være, at spørgsmål bør grupperes med samhørende emner som kriterium: "Do not switch from one topic to another" (Fink & Kosecoff, 1998:30). Ét argument til støtte for tematiske spørgeforløb er, at spørgsmålsrækkefølgen skal virke ”naturlig”. Problemet er imidlertid, at netop tematiske spørgeforløb undertiden bidrager til negative kontekstefekter (se fx Judd et al., 1991). Den slags utilsigtede måleproblemer lader sig muligvis forebygge ved at fravige det ovenfor nævnte kriterium. Yderligere en måde, hvorpå respondenters svarmotivation fremmes, er ved at placere de ”letteste” spørgsmål i begyndelsen af spørgeskemaet og henvise ”følsomme” spørgsmål til slutningen.

Endelig skal nævnes, at der heller ikke er konsensus om spørgeskemaers omfang og formatering. Nogle fremstillinger peger på, at spørgeskemaer ikke skal være ”for omfattende”, og at også formatering spiller en rolle, især i postspørgeskemaer (se fx Babbie, 1990; de Vaus, 2002; Fink & Kosecoff, 1998; Fowler, 1995; Frey & Oishi, 1995). Litteraturen

indeholder en lang række anvisninger, som interesserede læsere henvises til.

Kondenseringen af anbefalinger vedrørende formulering af ”gode” spørgsmål og udarbejdelse af ”gode” spørgeskemaer har underbygget følgende: ”No one has ever succeeded in compiling a perfect list of rules for writing good questions” (Salant & Dillman, 1994:91). Når afsættet tages i monosurveysk litteratur, hvis forfattere opfatter spørgeskema-konstruktion som ”håndværk” eller ”kunst”, synes litteraturen domineret af common sense og ”tommelfingerregler”: “... most questionnaire designers use a mixture of experience, intuition, judgement and common-sense ’rules’ to write what they believe will be clear, unambiguous questions” (Belson, 1981).

Blandt ”reglerne” findes fx råd om brug af ”hverdagssprog” samt præcise og korte spørgsmål. Spørgsmålet er, om sådanne ”regler” fører til ”gode” spørgsmål, dvs. til svar, der ikke er behæftet med måleproblemer. Mange regler er for diffuse til at kunne implementeres, hvortil kommer, at nogle regler synes gensidigt eksklusive. Når forfattere fx anbefaler ”præcise” spørgsmål kan det næppe imødegås, men derimod komme i konflikt med en samtidig anbefaling af korte spørgsmål. Om næppe altid umuligt er det ikke et enkelt forehavende at formulere præcise og korte spørgsmål.

### 3.3. MULTISURVEYSKE ANBEFALINGER

Efter at have præsenteret en række erfaringsbaserede anbefalinger om spørgeskemakonstruktion og spørgsmålsformulering i monosurveys, rettes opmærksomheden imod multisurveyske anbefalinger, dog først i det følgende afsnit, hvad oversættelse af spørgsmål angår. Da tendenserne er forskellige i mono- og multisurveysk litteratur, retter anbefalingerne sig kun delvist imod samme problematikker. Hertil kommer, at multisurveysk litteratur er af mere beskedent omfang end monosurveysk, hvorfor anbefalingerne er færre. Før specifikke anbefalinger præsenteres, er nogle indledende mere almene overvejelser om multisurveys hensigtsmæssige. De drejer sig om procedurer med henblik på udarbejdelse af spørgeskemaer og om forudsætninger herfor.

Også multisurveyske forfattere påpeger, at konstruktion af spørgeskemaer, der sigter imod sammenligning af nationer og/eller kulturer,

er en vanskelig og udfordrende opgave, såfremt krav om sammenlignelige data skal honoreres (se fx Johnson, 1997; Smith, i: Harkness et al., 2003). Selv om måleproblemer også forekommer i monosurveys, bliver problemerne forventeligt større, når grænser eller kulturer overskrides. Det gælder, uanset om forskere bringer allerede eksisterende spørgeskemaer i anvendelse i andre lande og/eller kulturer, eller om nye spørgeskemaer udarbejdes.

Når forskere involverer sig i udarbejdelse af tværsurveyiske spørgeskemaer, er en af de første afgørende beslutninger, om forhåndenværende spørgeskemaer skal anvendes, eller om nye bør udarbejdes. I førstnævnte tilfælde kommer eventuelle tilpasninger til ændrede nationale og/eller kulturelle kontekster på tale, mens den anden fremgangsmåde rejser spørgsmål om sekventiel, parallel eller simultan spørgeskemakonstruktion (se fx Behling & Law, 2000; Harkness et al., i: Harkness et al., 2003).

Mange forskere foretrækker eksisterende spørgeskemaer, eventuelt suppleret med påkrævede tilpasninger (se nedenfor). I tværnational surveyforskning er en ofte anvendt fremgangsmåde – undertiden kaldet ”stil samme spørgsmål”-proceduren – således også at tage afsæt i et nationalt udviklet spørgeskema og oversætte det med henblik på anvendelse i ét eller flere andre lande (se fx Harkness, 1999). I så henseende påpeger Harkness et al. væsentlige procedurale fordele: ”Not only does it make little sense to re-invent the survey wheel (...), but previous use is also seen as the best pretest. In addition, data are available on how questions performed” (i: Harkness et al., 2003:24). Nyere forskning peger imidlertid på, at umiddelbart overtagede og oversatte spørgsmål fordrer omhyggelige prætest, før de bringes i anvendelse i ændrede sammenhænge (op.cit.; se også Skinner, 2001).

Konstruktion af nye spørgeskemaer tilbyder flere muligheder end brug af eksisterende. Nye skemaer kan fx bringes til både at indeholde ”etiske” og ”emiske” spørgsmål (se nedenfor).<sup>19</sup> Når det handler om nye skemaer, synes sekventiel skemakonstruktion mest udbredt, dvs. at skemaet til at begynde med kun gælder for én national eller kulturel sammenhæng, hvorpå det ”eksporteres” til andre (se fx Harkness et al., 2003; Hoffmeyer-Zlotnik & Harkness, 2005). Sjældnere er parallel spør-

---

19. ”Etiske” spørgsmål er som nævnt kulturelt tværgående, mens ”emiske” er kulturelt specifikke (se afsnit 1.2.).

geskemakonstruktion, hvor skemaudkast (*source questionnaire*) udsættes for multikulturelle afprøvninger, fx i forbindelse med oversættelse. Hvad endelig arbejdskrævende simultan skemakonstruktion angår, følges formulering og oversættelse ad: "It is a technique which begins from a draft questionnaire in one language and produces final questionnaires in, usually, two languages through a process of paraphrase and translation between language one and language two" (Harkness et al., i: Harkness et al., 2003:32).

Flere forfattere gør opmærksom på, at konstruktion af multisurveyeske spørgeskemaer fordrer intensivt tværnationalt eller -kulturelt samarbejde og særlige kvalifikationer (se fx Smith, 2004). Samarbejde kan fx bidrage til at forebygge "forskningssimperialisme": "Research imperialism or safari research in which a research team from one culture develops a project and instrument and imposes it rigidly on other societies should be avoided" (Smith, i: Presser et al., 2004a:443). Teamsamarbejde efterlyses derfor ofte, fx af Kuechler (1987): "Conceptualization and questionnaire development should be a team effort. For each nation the research team should include at least one scholar with intimate knowledge of this nation (...). Ideally, each nation should be represented by at least one researcher native to this country and still residing there" (p. 327; se fx også Jiang et al., 2006).<sup>20</sup>

Også kulturel indsigt og sensitivitet er at foretrække (se fx Frey, 1970; Usunier, 1999; Word, 1992). Mange forskere synes uden videre at tage for givet, at de kender det nationale og/eller kulturelle univers, som en given multisurvey sætter fokus på: "An effort must be directed at getting as broad a spectrum of feelings about the topic so that questions are not arbitrarily narrowed down" (Word, 1992:35). Til opnåelse af kulturel viden og sensitivitet kan eksperter være nødvendige, enten ved at konsultere eller indlede samarbejde med dem: "Such teams would be asked to review the appropriateness of instrument content and data collection methods, and to identify other culture-specific considerations" (Johnson, 1998:13). Adskillige forfattere opfordrer til direkte samarbejde mellem surveyforskere og kulturelle eksperter (se fx Van de Vijver & Hamblen, 1996).

Uanset om multisurveyes baseres på "genbrug" eller nye spørgeskemaer, peger erfaringsbaseret metodelitteratur på vigtige problematik-

---

20. Vedr. eksempler på internationalt multisurveyesk samarbejde, se fx Smith, i: Harkness et al., 2003.

ker, der bør tages stilling til (se *oversigt 3.2*). Den første sætter fokus på, hvordan begreber og operationaliseringer gøres tværkulturelt kommensurable. Operationaliserende sammenlignelighed er imidlertid ikke tilstrækkelig. Hertil fordres også komparative tilpasninger over landegrænser og/eller kulturer. Endelig konfronteres forskere med semantiske udfordringer, når spørgsmål formuleres og svarkategorier udarbejdes: "It is still a standard believe in the social sciences that 'asking questions' is an art mastered by anyone, not requiring a specific methodology" (Mohler, 2005). Til udfordringerne hører oversættelsesproblemer (se afsnit 3.4.).

### Oversigt 3.2.

Aspekter vedrørende spørgeskemakonstruktion ifølge dele af multisurveysk metode-litteratur.

Begrebsoperationalisering	Spørgsmålstilpasning	Semantiske problemløsninger
Begrebssammenlignelighed	Love/regler	Fokusering
"Emisk" versus "etisk"	Politiske initiativer	Spørgsmåls længde
Antal empiriske indikatorer	Dateringer	Formulering
	Terminologier	Præciserende probes
	Betegnelser	Tidsrum
	Konventioner	Oversættelsesegnethed

Som i monosurveys er det at definere og operationalisere i multisurveyske problemstillinger indeholdte begreber en udfordring. Definitioner af abstrakte begreber – fx etnicitet, social integration og solidaritet – er som nævnt hverken sande eller falske, men mere eller mindre hensigtsmæssige, hvilket derfor også gælder begrebers operationalisering (se afsnit 3.2.). Operationalisering er at bevæge sig ned ad "abstraktionsstigen", så spørgsmål virker som indikatorer på mere eller mindre abstrakte begreber: "Problems associated with lack of conceptual equivalence across cultures have to do with the degree to which the concepts the researcher operationalizes in the source instrument ... exist in the same form in the thinking of members of the target culture" (Behling & Law, 2000:5). Dele af multisurveysk litteratur sætter fokus herpå og angår bl.a. dannelse af konceptuel sammenlignelighed, overvejelser vedrørende antallet af empiriske indikatorer samt forholdet mellem "emiske" og "etiske" spørgsmål.

Når landegrænser eller kulturer overskrides, øges risikoen for, at abstrakte begreber, som surveys baseres på, ikke bliver konceptuelt eller

strukturelt ækvivalente: "... the researcher needs to examine whether the construct has the same underlying meaning or significance in another country or context and hence the same measurement instrument can be used to effectively tap this construct in both contexts" (Douglas & Nijssen, 2002). Måske forekommer et givet begreb – det kan fx være national stolthed, solidaritet eller anomi – kun i nogle lande/kulturer, eller begrebet har ikke samme betydning i forskellige socio-kulturelle sammenhænge (se fx Bulmer, 1998; Okazaki & Sue, 1995). Ikke overraskende er det derfor et udbredt multisurveysk forlangende, at begreber og deres operationalisering skal generere sammenlignelighed (se fx Harkness et al., 2002; Johnson, 1998; Schaffer & Riordan, 2003; Smith, 2004; Smith & Bond, 1998). Selv om fordringen nyder almen anerkendelse, peger flere dog samtidig på, at surveyforskere sjældent lever op til den (se fx Smith, i: Presser et al., 2004a). Det skyldes måske bl.a., at der – som de følgende par sider anskueliggør – ikke er konsensus om anbefalinger, der følger i kravets kølvand.

Et i multisurveysk litteratur ofte nævnt begrebspar er "etisk" versus "emisk". Da ikke alle begreber er kulturelt universelle, er det som nævnt nødvendigt at tage stilling til, om multisurveyske begreber er af kulturuniversel "etisk" eller kulturspecifik "emisk" beskaffenhed (se afsnit 1.2.).<sup>21</sup> Adskillige forfattere gør den opfattelse gældende, at multisurveys i sagens natur udelukkende bør sætte fokus på "etiske" begreber og operationalisere dem på måder, der forebygger såkaldt *imposed etics*, dvs. at "... the procedures, experimental situations and measures used assume that the response being studied have the same meaning to the new participants as it did to those in the setting where the study was originally devised and that therefore responses will have an equivalent meaning in the new culture" (Smith & Bond, 1998:57; se fx også John-

---

21. "Emic and etic are terms used by some in the social sciences and the behavioral sciences to refer to two different kinds of data concerning human behavior. An "emic" account of behavior is a description of behavior in terms meaningful (consciously or unconsciously) to the actor or actress. An "etic" account is a description of a behavior in terms familiar to the observer. Scientists interested in the local construction of meaning, and local rules for behavior, will rely on emic accounts; scientists interested in facilitating comparative research and making universal claims will rely on etic accounts" (Wikipedia, The Free Encyclopedia).



son, 1998; Robert et al., 2006; Schaffer & Riordan, 2003). På den måde er der risiko for, at tværkulturelle sammenligninger bliver ”emiske”.<sup>22</sup>

Til forebyggelse af *imposed etics* anbefales undertiden et udvidet antal empiriske indikatorer til måling af samme begreb (se fx Braun & Scott, 1996; Smith, i: Harkness et al., 2003; Presser et al., 2004; Smith, i: Presser et al., 2004a; Van de Vijver & Leung, 1997). Hvis der er tvivl om, hvorvidt et abstrakt begreb er tværkulturelt relevant, lyder anbefalingen: *more is better*. Ved triangulerende overensstemmelse mellem flere empiriske indikatorer – gerne mindst tre, anfører Van de Vijver & Leung (1997) – menes data at blive mere komparativt robuste, mens der er grund til forsigtighed, hvis blot én indikator er statistisk ”afvigende” i forhold til de øvrige. Eller hvis temaet er religiøsitet, kan spørgsmål angående deltagelse i gudstjenester og andre ”ydre” aktiviteter være utilstrækkelige (se fx Harkness, 1999).

Atter andre plæderer for, at multisurveys bør baseres såvel på ”etiske” som ”emiske” begreber og begrebsoperationaliseringer, hvoraf følger, at spørgeskemaer, der ikke uden videre overtages fra andre sammenhænge, bliver hybrider både indeholdende generelle tværgående og kulturspecifikke spørgsmål: ”... we suggest a combined emic-etic or a derived etic approach when making cross-cultural comparisons” (Schaffer & Riordan, 2003:174; se fx også Berry, 1969; Johnson, 1998; Schmidt, i: Harkness et al., 2003; Triandis, 1992).<sup>23</sup> Én begrundelse nævnes af Usunier (1999), der er stærk tilhænger af kombinationstilgangen: ”... there are both areas of shared meaning and facets of meaning specific to each culture”.<sup>24</sup> En anden begrundelse påpeges af Smith: ”... the

---

22. *Imposed etics* er fx mulig, hvis religiøsitet alene måles ved overværelse af religiøse ceremonier, hvilket kan være problematisk i lande, hvor religiøsitet helt eller overvejende er et privat anliggende (se Jiang et al., 2006).

23. “... there are ideas in one language that have no direct or exact equivalents in another. If we are interested in measuring the etic construct, the use of emic instruments is an advantage, because it allows us to utilize the local terms, so we can tap directly into the cognitive system of the respondents” (Triandis, 1992:239).

24. Følgende citat fra Usunier (1999) anskueliggør tankegangen: “... the best way to generate insights as to conceptual equivalence is for a multilingual/multicultural team of researchers to use multiple source and target languages and to allow for decentering in the translation process (...). For this, researchers should start from a rather broad conceptual area which is almost etic because it corresponds to common problems in all languages/cultures. The procedure then consists in asking native speakers of the languages and cultures concerned to elicit key-words related to the broad conceptual area (...) as in cognitive mapping (...). A collaborative team of cross-cultural researchers then chooses the most frequently cited terms related to the concept in each language/culture and back-translates them; the team may then debrief to generate etic and emic

emic/etic approach indicates that sometimes one needs to do things differently in order to do them equivalently” (Smith, 2004:8)

Ifølge den nævnte opfattelse, der navnlig retter sig imod lande eller kulturer, der er markant forskellige, fordrer tværnationale eller kulturelle konceptuelle overlap, at overlap operationaliseres ved samme spørgsmål, hvorimod ikke-overlap kræver kulturelt specifikke spørgsmål: ”The emic/etic approach is useful if the common core is adequate for direct comparisons” (Smith, i: Presser et al., 2004a:446). Selv om flere forfattere plæderer for miks af ”etiske” og ”emiske” spørgsmål, men samtidig betoner behovet for særlig omhyggelig afprøvning af den slags spørgeskemaer, er anbefalingen som de foregående stort set uden forskningsmæssig grundthed (se fx Berry, 1999; Harkness et al., i Harkness et al., 2003) og derfor placeret i dette kapitel.

Hvad enten multisurveyske spørgeskemaer ”genbruges” eller ej, og uanset det dermed forbundne mindre eller større arbejde med at definere og operationalisere begreber, er spørgsmålstilpasning til forskellige nationale og/eller kulturelle sammenhænge påkrævet (se *oversigt 3.2*). Tilpasninger, der først og fremmest sigter imod dannelse af funktionel ækvivalens, kommer også på tale ved gennemførelse af monosurveys, men tilpasningsbehovet menes større i multisurveys (se fx Behling & Law, 2000; Harkness et al., i: Harkness et al., 2003; Johnson et al., 2001; Smith, 2004; Smith, i: Presser et al., 2004a). Funktionelle forskelle, der fordrer spørgsmålstilpasning, retter sig imod fravær af ”ækvivalente objekter”, fx vekslende love eller andre regelsæt, politiske initiativer og programmer, dateringer, terminologier og betegnelser.

Hvad spørgsmålstilpasning indebærer, kan eksemplificeres: I nogle lande findes fx en præsident, men ikke en ombudsmand, mens det forholder sig omvendt i andre. Parlamentsvalg og andre offentlige valg-handlinger finder almindeligvis ikke sted på samme tidspunkt i forskellige lande, ligesom tidspunktet for afsluttet folkeskole eller gymnasium kan variere. Selv et tilsyneladende simpelt spørgsmål om ægteskabelig stilling kan angå et ”ikke-ækvivalent objekt”, såfremt en multisurvey også omhandler lande eller kulturer, hvor mænd tilladeligt kan være polygame (Smith, 2004). Endelig henviser betegnelserne *states*, *provinces* og *länder* langt fra til ækvivalente objekter. Der er med andre ord behov for såvel faktuelle som terminologiske tilpasninger, men også konventionsstyrede

---

conceptual dimensions. The frequency of citation of certain facets enables to assess the degree of emicity”.

tilpasninger kan komme på tale, fx læsning af spørgsmål fra højre til venstre (se fx Johnson et al., 2001).

Monosurveysk metodelitteratur indeholder en lang række anbefalinger vedrørende semantiske problemløsninger (se afsnit 3.2.), mens multisurveysk litteratur – bortset fra oversættelse af spørgsmål – er væsentlig mindre ”talende” og i høj grad lægger sig op ad erfaringsbaserede anbefalinger fra monosurveysk litteratur (se fx Johnson, 1998; Johnson et al., i: Hoffmeyer-Zlotnik & Harkness, 2005; Robert et al., 2006; Smith, i: Presser et al., 2004a; Van de Vijver & Leung, 1997). Den gennemgående tendens er derfor, at multisurveysk metodelitteratur ikke hidtil har ydet selvstændige bidrag, hvad spørgsmålsformulering angår (se *oversigt 3.2.*).

Derimod peges der i en del af multisurveysk metodelitteratur på erfaringsbaserede anbefalinger, der kendes fra monosurveysk litteratur (se *oversigt 3.1.*). Én anbefaling handler om spørgsmåls éndimensionelle fokusering, dvs. at spørgsmål kun bør rette sig imod afgivelse af én information (se fx Smith, i: Presser et al., 2004a:445). En anden velkendt anbefaling angår spørgsmåls længde, hvor flere kræver *brevity*: ”Use short, simple sentences of less than sixteen words” (Johnson, 1998:15). Hertil kommer en gentagelse af opfordringen til størst mulig sproglig enkelhed, klarhed og éntydighed, herunder advarsler imod slang og metaforer (se fx McKay, 1996) samt fravalg af vage og tvetydige ord, fx abstrakte ord og vage kvantificerende adverbier (se fx Johnson, 1998; Robert et al., 2006; Scheuch, 1993).<sup>25</sup> Enkelte forfattere fremhæver tillige behovet for at definere og/eller eksemplificere uundgåelige tvetydige ord, ligesom nødvendigheden af klare tidsmæssige afgrænsninger påpeges (se fx Smith, i: Presser et al., 2004a).

Én anbefaling, som flere nævner, er dog ikke en gentagelse i forhold til monosurveysk litteratur, nemlig at spørgeskemaer bør udarbejdes, så de er egnede til oversættelse. Fx Van de Vijver og Leung (1997) påpeger, at ikke alle tekster – fx spørgeskemaer – er lige oversættelige: ”A text is poorly translatable when the loss of salient characteristics cannot be avoided in translation” (p. 38). Alene af hensyn til senere oversættelse bør spørgsmål, mener Behling og Law (2000), være så præcise som

---

25. I litteraturen nævnes lykke som ét eksempel på abstrakte ord (se fx Smith, i: Presser et al., 2004a). Kvantificerende adverbier er fx *ofte* eller *en gang imellem*.

muligt, kortest mulige, uden metaforer og med behørlige sproglige præciseringer (*probes*).

### 3.4. OVERSÆTTELSE

Modsat monosurveysk litteratur fokuserer megen multisurveysk i sagens natur på oversættelse af spørgeskemaer. Det drejer sig som regel om oversættelse af eksisterende masterskemaer (se fx Harkness et al., i: Presser et al., 2004). Enkelte titler angår udelukkende oversættelse (se fx Behling & Law, 2000; Forsyth, 2006; Harkness et al., 2004; Weidmer, 1994), men i de fleste tilfælde tages oversættelse under behandling sammen med andre problematikker, fx spørgeskemakonstruktion (se fx litteraturoversigt i: Harkness & Schoua-Glusberg, 1998). Oversættelse kommer ikke kun på tale i tværnationale surveys, da også nationale kan fordre oversættelse. Det gælder i lande som Belgien, Schweiz, USA og Canada (se fx Presser et al., 2004). I USA vil der fx ofte være behov for at oversætte spørgsmål til spansk. Selv i Danmark kan der være behov for oversættelse, fordi ikke alle respondenter behersker det danske sprog.

Den righoldige litteratur om oversættelse af spørgeskemaer er overvejende erfaringsbaseret og kondenseres derfor her: "Perhaps no aspect of cross-national survey research has been less subjected to systematic, empirical investigation than translation" (Smith, 2004). Tilsyneladende findes der stort set ingen systematiske eksperimenter, som anbefalinger baseres på: "... development of scientifically based translation has languished" (Smith, i: Presser et al., 2004a:447). At feltet er udforsket kan fx skyldes, at mange forskere uden videre mener at kunne oversætte spørgeskemaer: "Similar to the believe about asking questions, scientists believe in their intrinsic ability to translate items" (Mohler, 2005). Såfremt oversættelse anskues som et rent "teknisk" anliggende, øger det risikoen for data, der ikke er sammenlignelige. Oversættelse af spørgeskemaer er ikke "... a technical matter which has little bearing on the outcome. It is of epistemological consequence as it influences what is found" (Temple, 1997:614; se fx også Birbili, 2000; Bullinger, 2003).

Som det vil fremgå, er oversættelseslitteraturen ingenlunde konsensuel. Den retter sig navnlig imod almene ledetråde for oversættelse samt anbefalinger vedrørende oversættelsesprocedurer, herunder fx aktører og kvalifikationer. Hvordan oversættelser præstest, er også en ofte

omtalt problematik (se kapitel 5). Muligvis medvirker litteraturens sammensathed og manglende forskningsbaseret til at forklare, at der også i praksis er uenighed om, hvordan oversættelser bedst gennemføres: "... many cross-national studies do not insist on any systematic method of translating the source questionnaire into other languages" (O'Shea et al., 2005)<sup>26</sup> (se *oversigt 3.3.*).

Ud over at spørgeskemaer bør konstrueres og formuleres, så de er egnede til oversættelse (se afsnit 3.3.), ligger det ligefor, at den mest overordnede, men dog vage ledetråd er almen dataækvivalens på tværs af lande eller kulturer (se fx Behling & Law, 2000; Harkness, 1999; Herdman et al., 1996; Johnson, 1998). Derfor er også semantisk ækvivalens vigtig: "In survey research, questionnaire translations are generally, if vaguely, required to ask the same questions and offer the same response options as a source text. Rightly or wrongly, they are expected to do so by means of a close rendering of the source questionnaire" (Harkness & Schoua-Glusberg, 1998:92; se fx også: Jowell, 1998). Hvad semantisk ækvivalens er, og hvordan den tilsikres i surveypraksis, er der imidlertid ikke konsensus om.

Grundtanken er, at såvel spørgsmål som svarkategorier så vidt muligt bør "bevare" deres oprindelige betydning efter afsluttet oversættelse (se afsnit 2.3.). At dette kan være et vanskeligt forehavende – i mange tilfælde grænsende til umuligt – er formodentlig selvindlysende for de fleste sprogforskere og tillige for en del multisurveyske forfattere, fx Harkness og Schoua-Glusberg (1998): "Languages are not isomorphic and so translation cannot be expected to operate on a one-to-one basis across languages. This means that what goes in (the source language text) cannot be completely matched by what comes out (the target language text). Indeed, a mechanistic notion of input and output is itself misleading" (p. 93). Det er derfor ikke overraskende, at nogle forfattere opfatter oversættelse af spørgsmål som ét af de svageste led i en proces, der gerne skulle generere sammenlignelige data (se fx Smith, i: Presser et al., 2004a).

---

26. Angående EVS og ISSP skriver O'Shea et al. (2005): "In the 1999/2000 EVS ... the majority of participating countries did not check their translated questionnaires by back-translating, the only method specified (...). And in the ISSP, similar omissions are common (...). The effects of these omissions are often serious. For instance, analysis of the 1999/2000 EVS has highlighted problems of equivalence in respect of several questions in several countries".

I forlængelse af fordringen om semantisk ækvivalens har andre generelle ledetråde at gøre med lingvistikens tre hovedområder, nemlig grammatik, semantik og pragmatik, hvor litteraturens fokus i særlig grad er på semantiske refleksioner. Det er velkendt, at sætningers grammatiske struktur kan påvirke deres betydning, men hvordan en af oversættelse nødvendiggjort ændret grammatisk struktur påvirker spørgsmåls betydning, er der kun få overvejelser om og eksempler på i multisurveysk litteratur (se fx Birbili, 2000; Ercikan, 1998; Harkness, i: Harkness et al., 2003).

### Oversigt 3.3.

Anbefalinger vedrørende oversættelse af spørgeskemaer ifølge multisurveysk litteratur.

Generelle ledetråde	Oversættelsesprocedurer
Oversættelseegnethed	Påbegyndelse
Semantisk ækvivalens	Aktører
Grammatiske ledetråde	Kvalifikationer
Semantiske ledetråde	Opgavebeskrivelser
Pragmatiske ledetråde	Dokumentation
	Samarbejde/teams
	Procedureeksempler
	Prætest (kapitel 5)

Alene fraværet af grammatiske "ækvivalenter" i sprog, som spørgeskemaer oversættes til, kan føre til måleproblemer, mener Ercikan (1998), der særligt påpeger vanskelighederne ved at "transportere" syntaks fra ét sprog til et andet: "Syntactical style is one of the most difficult features to carry over from one language to another" (p. 544). Visse grammatiske strukturer er resistente over for umiddelbar transformation til andre sprog med informationstab til følge (op.cit.). Også Harkness anstiller grammatiske betragtninger og diskuterer fx den rolle, som grammatisk køn spiller med oversættelse af spørgsmål fra engelsk til tysk: "People have sexes; languages may or may not have grammatical gender. Grammatical gender assigns animate and inanimate objects a gender" (Harkness, i: Harkness et al., 2003:49). Nogle sprog kendetegnes ved komplekst grammatisk køn (fx spansk), mens andre er enklere (fx engelsk) eller helt fraværende (ungarsk).

Selv om grammatiske forskelle mellem spørgeskemaer, der foreligger på to eller flere sprog, kan påvirke spørgsmåls semantik og dermed

datas sammenlignelighed, er multisurveyske forfatteres bevågenhed langt større, når det handler om ords og sætningers vekslende betydning på forskellige sprog: "Lack of semantic equivalence across languages is our label for problems that hamper efforts to identify words and phrases in the target language that have meanings matching those in the source language instrument" (Behling & Law, 2000:4). Det drejer sig både om oversatte ords, spørgsmåls og svarkategoriens semantik (se fx Behling & Law, 2000; Birbili, 2000; Braun & Harkness, i: Hoffmeyer-Zlotnik & Harkness, 2005; Ercikan, 1998; Harkness, i: Harkness et al., 2003; Harkness & Schoua-Glusberg, 1998; Heath et al., 2005; Smith, i: Harkness et al., 2003).

Valg af enkeltord, der indgår i spørgsmål og svarkategorier, rejser mindst to problematikker: Findes det pågældende ord på det sprog, som et spørgeskema skal oversættes til? Har ordet i givet fald samme (leksikalske) betydning på begge sprog? Mens det første spørgsmål næsten altid vil blive besvaret bekræftende, forholder det sig modsat med det andet. I litteraturen nævnes enkelte eksempler på ord, der vanskeligt eller ikke kan oversættes: "In some cases, a term may be used for which there is no direct equivalent in another language, for example *philotimo* in Greek or 'face' in Chinese. Equally, colloquial terms need to be translated into a colloquial equivalent" (Douglas & Nijssen, 2002).<sup>27</sup> Yderligere et eksempel er det japanske *giri*, der angår ære og social forpligtelse, og som ikke kan oversættes til vestlige sprog (se fx Sasaki, 1995; Smith, i: Presser et al., 2004a).<sup>28</sup> Et sidste eksempel er fra Harpaz (2000): "... in

---

27. "According to Triandis and Vassiliou (1972, pp. 308-309) someone who is *philotimos* "behaves towards members of his in-group the way they expect him to behave." They write: "A person who has this characteristic is polite, virtuous, reliable, proud, has a 'good soul,' behaves correctly, meets his obligations, does his duty, is truthful, generous, self-sacrificing, tactful, respectful, and grateful." When a large proportion of the Greeks report that they see themselves as *philotimos* there are two basic orientations in the interpretation of this finding. First, there is something to *philotimo* that the Greeks have and "we" do not have. Second, *philotimo* (*philotimos* behavior) can be observed with us, it is just not explicit in our language or self-reports; for example, it is a particular blend of common human dispositions that has become salient in the Greek language and society probably for socio-cultural-historical reasons" ([http://ebooks.iaccp.org/ongoing\\_themes/chapters/poortinga/poortinga.php?file=poortinga&output=screen](http://ebooks.iaccp.org/ongoing_themes/chapters/poortinga/poortinga.php?file=poortinga&output=screen)).

28. "There is no possible English equivalent and of all the strange categories of moral obligations which anthropologists find in the cultures of the world, (*giri*) is one of the most curious (...). Despite the rapid change of Japanese life, ideas and society, the "curious" concept of "*giri*" has remained and still strongly governs Japanese social behavior. To analyze "*giri*" Berdict's famous work, *The Chrysanthemum and the Sword*, might be a useful starting point. Although this book is now a "classic", it is full of errors and misunderstandings, such as the confusion of "*giri*" with

French, there is no word for 'achievement', while in Japanese there seems to be no word for 'decision making'" (Harpaz, 2000).

Ved oversættelse af spørgeskemaer er det imidlertid en langt større udfordring, at "samme" ord ofte får en anden betydning, når det oversættes, fx fordi ordet har socio-kulturelle konnotationer: "Some concepts may not have equivalents in another language, and even when they do, the exact meaning may be utterly different" (Harpaz, 2000). Dette er velkendt for sprogforskere, men kun i fokus blandt fåtallige multisurveyske forfattere, fx Smith (2004), der noterer: "Even cognates between closely related languages can substantially differ" (p. 4; se fx også Birbili, 2000; Smith, i: Presser et al., 2004a; Usunier, 1999). Nogle forfattere anbefaler derfor overskridelse af leksikalsk semantik: "On those occasions where two languages do not offer direct lexical equivalence several researchers and linguists suggest that one's efforts should be directed 'towards obtaining conceptual equivalence without concern for lexical comparability'" (Birbili, 2000).

Den sporadiske litteratur om ændrede ordbetydninger har overvejende karakter af eksempelsamling. Som ét eksempel nævner Smith (2004) *equality* i forhold til *égalité*, der ikke betyder det samme i USA, det engelsktalende Canada og det fransktalende Canada. Tilsvarende gælder *educación* i forhold til *education*, der har forskellig betydning i spansk- henholdsvis engelsktalende områder i USA (Greenfield, 1997). Yderligere et eksempel hidrører fra Jiang et al. (2006), der påpeger, at *violence* udelukkende er *brute force*, når ordet oversættes til kinesisk, mens ordets betydningsvidde er bredere på engelsk.<sup>29</sup>

Den omstændighed, at ord ofte får anden betydning, når de oversættes, er en endnu større udfordring end eksemplerne antyder, da

---

"chuu". Thus we have to take care to avoid citing such errors from her work. The concept of "giri" is even now accepted as forming an important part of Japanese social relationships and has been a perpetual theme in a variety of arts, such as the plays of Chikamatsu Monzaemon, the Edo Playwright, in ningyō-jōruri (puppet dramas) and Kabuki. Even modern television soap operas and cinema films use "giri" themes to draw tears from the audience. What then is "giri"? A general definition may be "duty" or "obligation" which arises from a social interaction with another person. But this fails to reveal a wide range of significant nuances (<http://academic.csuohio.edu/makelaa/history/courses/his373/giri.html>).

29. Shenkar & von Glinow (1994) nævner, at et ord som *autonomy*, der er et centralt organisationsteoretisk begreb, "... cannot be adequately translated into Chinese, or that alternate Chinese terms, such as "right of self-determination" (zi zhu quan) convey a quite different meaning" (s. 67). Yderligere et eksempel: "There are ... some indications that the concept of "commitment," which receives much attention in the management literature, may not be applied to non-Western cultures" (Harpaz, 2000).



også svarkategorier oversættes. At ”transportere” svarkategoriens semantik fra ét sprog til et andet kan være vanskeligt, hvilket fx påpeges af Harkness: ”... we cannot expect to find neatly corresponding structural or semantic components across languages” (i: Harkness et al., 2003:56). Betyder *often* og *now and then* fx det samme på fransk? Eller hvad betyder *strongly agree*, *agree*, *neither agree nor disagree*, *disagree* og *strongly disagree* på fransk?

Hverken i spørgeskemaer eller i andre tekster optræder ord almindeligvis alenestående, men betydningsdanner sætninger, fx survey-spørgsmål. Dele af litteraturen omhandler denne problematik med særlig fokus på grader af oversættelsestekstnærhed (se fx Braun, i: Harkness et al., 2003; Harkness, i: Harkness et al., 2003; Harkness & Schoua-Glusberg, 1998; Smith, 1998; Van de Vijver & Hamblen, 1996; Weidmer et al., 1999). Mens nogle forfattere – det er som regel fortalere for stimulus-respons-modellen, der forudsætter standardiseret dataindsamling – anbefaler ”tæt” oversættelse sigtende imod, at ”samme spørgsmål” stilles til alle respondenter, er andre forbeholdne, fordi ”tæt” oversættelse kan være så tekstnær, at formuleringer bliver stive, unaturlige og svært forståelige med den virkning, at data bliver usammenlignelige: ”Translated versions that are clear and make sense to respondents may not be as close as researchers might like” (Harkness, i: Harkness et al., 2003:47). Eller med en tilsvarende formulering: ”In order to convey the same meaning in different countries, details in the formulation of a question may require changing” (Braun, i: Harkness et al., 2003:67). For sådanne forfattere er betydningsækvivalens vigtigere end ”bogstavelig” ækvivalens (se fx også Jiang et al., 2006; Sasaki, 1995).<sup>30</sup>

Det kan umiddelbart virke ukompliceret at oversætte spørgeskemaers ”enkle” formuleringer: ”Survey translation looks rather simple” (Harkness, i: Harkness et al., 2003:43). Men spørgeskemaer er vanskelige at oversætte. I så henseende er omhyggeligt arbejde ikke en tilstrækkelig forudsætning. Som eksempel på manglende omhu nævner Heath et al. (2005) et spørgsmål fra Eurobarometer: *Are you close to a political party?*, der ved tilbageoversættelse fra den franske version bliver til *Are you closer to one party than the others?* I et andet eksempel, der viser, hvad tekstnær

---

30. Se også følgende citat vedrørende sprogs kulturelle følsomhed: ”... without contextual grounding, questions which rely solely on literal translation both the questions and the responses are likely to be faulty” (Sasaki, 1995:109).

oversættelse kan forvolde af problemer, bliver *He looks for adventures* til *Il cherche les aventures*, hvorved franske mænd spørges om noget andet, nemlig seksuelle eventyr (Harkness et al., i: Presser et al., 2004a:462). Yderligere et fældeeksempel er, når *out of sight, out of mind* ved tilbageoversættelse bliver til *blind and demented* (Jowell, 1998).<sup>31</sup>

Med de sidste eksempler er pragmatikkens relevans antydet, hvilket i denne sammenhæng vil sige, hvordan spørgsmål bruges og kommunikativt tildeles mening af respondenter. Det er en problematik, som enkelte forfattere – også tidligere nævnte – sætter fokus på og eksemplificerer (se fx Braun & Harkness, i: Hoffmeyer-Zlotnik & Harkness, 2005; Clark & Schober, 1992; Harkness, i: Harkness et al., 2003; Katan, 1999). Harkness betoner i flere sammenhænge behovet for pragmatiske overvejelser, når spørgsmål oversættes, dog uden at opstille operationelle ledetråde herfor: "... survey translations that focus on conveying lexical meaning and exclude consideration of pragmatic meaning have an Achilles heel. Pragmatic meaning ... gives contextual considerations more emphasis. Contextual considerations are cultural considerations" (i: Harkness et al., 2003:48).

Ifølge Harkness indebærer pragmatiske betragtninger – overvejelser om spørgsmål og svarkategorier i brug – at rette opmærksomhed imod kulturelt betingede betydninger, men hun understreger, at det er en problematik, som kun få surveyforskere indtil videre har sat fokus på, fx Smith: "... the context can alter question meaning even though questions in one survey are essentially the same in format and semantics as those in another survey" (i: Harkness et al., 2003:71).<sup>32</sup> Hvis Harkness' anbefaling tages ad notam, er "gode" oversættelser ikke længere nok: "Poor translations of good questions mean respondents read and respond to a question they should not have been asked (...). However, technically well-translated questions that are understood differently in different cultures are equally problematic" (Braun & Harkness, i: Hoffmeyer-Zlotnik & Harkness, 2005:104). Følgelig mener Braun og

---

31. Endnu et eksempel angår *feeling blue*, der er vanskelig at oversætte, fordi farven ikke har samme betydning i forskellige kulturer og på forskellige sprog (Harkness, i: Cohen & Lepkowski, 2004:88).

32. Smith nævner et interessant eksempel på, hvordan "samme" spørgsmål om adgang til fri abort kan få vidt forskellige kulturelt betingede betydninger i USA og Costa Rica (Smith, i: Harkness et al., 2003).

Harkness, at der er behov for udvikling af nye procedurer til kontrol af oversættelser (se kapitel 5).

Med ovenstående er litteraturens generelle ledetråde for oversættelse af spørgsmål introduceret, hvorpå fremstillingen bevæger sig i retning af oversættelsesprocedurer, hvorom der heller ikke er konsensus (se *oversigt 3.3.*). Hvornår bør oversættelse påbegyndes og med deltagelse af hvilke aktører? Spørgsmålet har mere end ét svar. I overensstemmelse med dominerende praksis tages normalt for givet, at oversættelse først forberedes og gennemføres, når et masterskema foreligger. Enkelte anbefaler imidlertid, at oversættelse om muligt integreres i selve undersøgelsesdesignet og dermed undgår at fungere som ”teknisk” tilføjelse: ”Researchers should ... make translations an integrated part of the development of studies” (Smith, i: Presser et al., 2004a:450; se også Harkness, i: Harkness et al., 2003).

Uanset om oversættelse virker som appendiks eller ej, kommer mindst tre typer af aktører på tale: (1) oversættere, (2) reviewere (*translation reviewers*) og (3) ”dommere” (*translation adjudicators*) (se fx Behling & Law, 2000; Harkness, i: op.cit.; Harkness & Schoua-Glusberg, 1998; McKay et al., 1996). Med muligt personsammenfald udfører oversættere det praktiske arbejde med at oversætte, reviewere bedømmer arbejdsresultatet, og ”dommere” træffer endelig afgørelse vedrørende valg af oversættelse, evt. med bistand af en fjerde aktørtype, nemlig konsulenter med særlig ekspertise. Det er fx den procedure, der anvendes af *European Social Survey* (ESS).

For at kunne virke som oversætter eller på anden måde være aktør i oversættelsesprocessen fordres bredspektrede kvalifikationer. Med overvejelserne om almene ledetråde for oversættelse in mente, er det ikke overraskende, at der heller ikke er konsensus om, hvad nødvendige kvalifikationer egentlig er (se fx Birbili, 2000; Harkness, i: Harkness et al., 2003; Harkness & Schoua-Glusberg, 1998; Hoffmeyer-Zlotnik, 2004; McKay et al., 1996). Hensigten er at oversætte spørgsmål, så de forbliver de ”samme” på andre sprog, men det er som nævnt en vanskelig – undertiden umulig – opgave. Ud over det indlysende, at oversættere mindst bør beherske to – gerne flere – sprog, nævnes følgende kvalifikationer til fremme af ”gode” oversættelser: oversættelseserfaring, erfaring med spørgeskemakonstruktion, kulturel ekspertise og konkret viden om den pågældende multisurvey.

Hvad oversættelseserfaring angår, betoner fx Harkness (i: Harkness et al., 2003), at erfaringerne også bør omfatte oversættelse af de helt særlige tekster, der udgøres af spørgsmål og svarkategorier. Det er tekster, som selv erfarne oversættere ofte ikke har prøvet at oversætte. Det er derfor nærliggende også at kræve erfaring med spørgeskemakonstruktion eller i det mindste viden herom. Surveyrelevant kulturel ekspertise nævnes som yderligere en fornøden kvalifikation: ”The problem ... is a problem of lack of knowledge, sometimes of ignorance concerning cultural differences. It is problematic that our translators are bilingual, but not bicultural” (Hoffmeyer-Zlotnik, 2004:3). Ikke kun multisurveyske forskere, men også oversættere bør, påpeger Birbili (2000), have tilegnet sig *'intimate' knowledge of the culture* (se også Frey, 1970). Endelig fremhæver enkelte forfattere, at oversættere bør have konkret viden om den multisurvey, hvis spørgeskema oversættes, fx med hensyn til problemstillinger og design.

Uanset kvalifikationer kan oversættere ikke udføre deres arbejde, hvis arbejdsopgaver er vage eller ubeskrevne. Opgavebeskrivelser er derfor yderligere en forudsætning for oversættelse af spørgeskemaer (se fx Harkness, i: Harkness et al., 2003; Harkness & Schoua-Glusberg, 1998; Smith, i: Presser et al., 2004a). Ifølge disse og andre forfattere har oversættere brug for præcise skriftlige opgavebeskrivelser. Den slags beskrivelser aflaster oversætteres arbejde og muliggør senere vurderinger af, om de for en oversættelse gældende *task specifications* er efterkommet.

Opgavebeskrivelser er ét aspekt af en bredere anbefaling, som enkelte nævner, nemlig dokumentation af den samlede oversættelsesproces, bl.a. med sigte på at skabe åbenhed om trufne beslutninger (se fx Harkness, i: Harkness et al., 2003; Harkness et al., i: Presser et al., 2004a; Harkness & Schoua-Glusberg, 1998; Harpaz, 2000). Harkness beklager dog samtidig, at multisurveyske oversættelser sjældent dokumenteres i tilstrækkeligt omfang (i: Harkness et al., 2003). Ifølge Harkness handler dokumentation navnlig om diverse baggrundsdokumentation til oversættere, dokumentation af selve oversættelsesprocessen samt dokumentation af det endelige produkt: ”In a multiple language context, documentation is as critical to the translation process as it is to the final product” (Harkness et al., op.cit.:470).

Andre udvider anbefalingen af dokumentation til at omhandle samtlige undersøgelsesmæssige procedurer, beslutninger etc. (se fx Jowell, 1998; Smith, i: Presser et al., 2004a). Dokumentation menes fx at

fremme datasammenlignelighed: "Solid documentation is more than just a good practice that facilitates primary and secondary analysis. It enhances comparability from the start by forcing researchers to detail what procedures are being used in each country and how comparable they are" (Smith, op.cit.:452). Hertil kommer, at dokumentation gør det muligt at "kigge forskere over skulderen" (se fx Jowell, 1998).

Oversættere og andre aktører kan indgå i samarbejdsrelationer, hvis art bl.a. afhænger af de i *oversigt 3.4.* nævnte procedureeksempler (se fx Behling & Law, 2000; Bullinger, 2003; Forsyth, 2006; Harkness, 1999; Harkness et al., i: Presser et al., 2004a; McKay et al., 1996; Rode, 2005; Smith, 2004; Usunier, 1999). Udover oversættere, der arbejder alene og uden videre leverer deres skriftlige produkt, nævner litteraturen forskellige former for samarbejde, hvis fordele og ulemper der er uenighed om. Det gælder fx samarbejde mellem oversættere, oversættere samarbejdende med forskere, oversættere både samarbejdende med forskere og konsulenter med særlig ekspertise samt egentligt komité- eller teamsamarbejde. Kvalitetssikrende holdarbejde på tværs af grænser og/eller kulturer anbefales af flere forfattere, fx af Harkness et al. (i: Presser et al., 2004a): "Team approaches to survey translation and assessment have been found to provide a richer pool of options to choose from translating items and a balanced critique of versions" (p. 464).<sup>33</sup> Komité- eller teamsamarbejde kan forebygge *research imperialism* og *safari research* (se fx Smith, i: Presser et al., 2004a).

Bedre og mindre idiosynkratiske oversættelser synes især mulige, påpeger Harkness (op.cit.), når teams både har sproglig og kulturel kompetence.<sup>34</sup> Til sikring af oversættelsers kvalitet anbefaler Harkness og andre derfor et udstrakt samarbejde mellem relevante fagdiscipliner, fx med sprogforskere: "Linguistics can be expected to offer an ongoing

---

33. I forlængelse af anbefalingen af teamsamarbejde udvikler Harkness en efter hendes opfattelse anbefalelsesværdig teammodel kaldet TRAPD med fem stadier: (1) oversættelse (*translation*), (2) review, (3) afgørelse (*adjudication*), (4) prætest (*pretesting*) og (5) dokumentation (*documentation*) (se fx Harkness, i: Harkness et al., 2003). Modellen kan anvendes på flere måder, fx ved "paralleloversættelse" (se nedenfor).

34. Yderligere et citat illustrerer anbefalingen af samarbejdende teams: "The team should bring together the mix of skills and disciplinary expertise needed to decide on optimal versions. Collectively, members of this team must supply knowledge of the study, of questionnaire design, of fielding processes (...). The team is also required to have the cultural and linguistic knowledge needed to translate appropriately in the required varieties of the target language" ([http://naticent02.uuhost.uk.uu.net/ess\\_docs/R3/Methodology/r3\\_translation\\_guidelines.doc](http://naticent02.uuhost.uk.uu.net/ess_docs/R3/Methodology/r3_translation_guidelines.doc)).

refinement of our understanding of different types of meaning” (Harkness et al., i: Presser et al., 2004a:472). Det gælder også kommunikationsforskere med kendskab til forhold, der fører til tværkulturelle vildfarelser.

### Oversigt 3.4.

#### Oversættelsesprocedurer.

---

- Simpel direkte oversættelse
  - Modifieret direkte oversættelse
  - Tilbageoversættelse
  - Decentrering
  - Paralleloversættelse
  - Komitéoversættelse
- 

Samarbejdsrelationer veksler fra multisurvey til multisurvey og afhænger ikke mindst af den valgte oversættelsesprocedure, som litteraturen nævner vekslende – mere eller mindre konkurrerende – muligheder for (se fx Behling & Law, 2000; Birbili, 2000; Harkness, i: Harkness et al., 2003; Harkness & Schoua-Glusberg, 1998; Harpaz, 2000; McKay et al., 1996; Sperber et al., 1994; Smith, 2004; Smith, i: Presser et al., 2004a). Nogle ofte omtalte procedurale eksempler præsenteres i det følgende (se *oversigt 3.4.*). Det er som nævnt umuligt at vælge mellem procedurerne på forskningsmæssigt grundlag, ligesom nogle af dem ikke udelukker hinanden.

Direkte oversættelse, som langt de fleste forfattere forbigår eller er forbeholdne overfor, findes i to versioner: simpel eller modifieret. Simpel direkte oversættelse – undertiden kaldet *one-for-one translation* (se fx Harkness & Schoua-Glusberg, 1998) – er udbredt i surveypraksis og indebærer, at et spørgeskema videregives til en oversætter, der udfører arbejdet og returnerer skemaet. Proceduren er således både fordringsløs og omkostningslet, men som regel ikke anbefalelsesværdig (se fx Smith, i: Presser et al., 2004a; Sperber et al., 1994). Når direkte oversættelse modificeres, er det ensbetydende med, at oversættere mødes et par gange med eksperter eller et ekspertpanel. Ifølge Behling & Law (2000) er proceduren mere informativ end simpel direkte oversættelse, og spørgeskemaer samtidig lettere at kvalitetsvurdere. Det er dog kun få forfattere, der nævner og anbefaler modifieret direkte oversættelse.

Derimod nævnes tilbageoversættelse af adskillige, dog ofte uden at blive anbefalet: ”This is a procedure whereby the questionnaires ... are

first translated into a form that is locally understandable, and then translated back into the first language by someone independent of the study, to check that there have been no mistranslations” (Smith & Bond, 1998:33; se fx også Ercikan, 1998). Proceduren, der stadig anvendes i praksis, og som kan kombineres med andre procedurer, fx komitéoversættelse, omfatter normalt fire faser: oversættelse, en anden oversætters tilbageoversættelse til det oprindelige sprog, sammenligning af de to oversættelser og udarbejdelse af en revideret oversættelse. Processen kan principielt fortsætte, indtil alle ”uoversstemmelser” mellem oversættelse og tilbageoversættelse er elimineret.

Mange har i årevis advokeret for (se fx Brislin, 1970) og anbefaler fortsat tilbageoversættelse: ”Translation/back-translation has substantial source language transparency”, mener Behling & Law (2000:21), og peger på de to versioners umiddelbare sammenlignelighed og kvalitets-sikrende brugbarhed. De kritiske røster, der også er af ældre dato (se fx Alwin et al., 2002),<sup>35</sup> bliver dog hyppigere og anfægter procedurens præmis: ”Back translation operates on the premise that ‘what goes in ought to come out’ ”(Harkness, i: Harkness et al., 2003:41). Selv om proceduren selvsagt kan afdække nogle problemer i forbindelse med oversættelse, har den en basal mangel: ”No assessment is made directly of the adequacy of the target language questions (...), so a poorly worded item that back-translates successfully goes undetected” (Smith, i: Presser et al., 2004a:446).

Proceduren – en *black box* procedure – kan derfor indgyde ”falsk tryghed”: ”The process of back translation ensures literal accuracy, and facilitates the recognition of mistakes; yet it may lead the investigator to assume that equivalence has been achieved when it has not” (Harpaz, 2000). Hvis involverede forskere tilmed kun har kendskab til ét eller et par af de sprog, som et spørgeskema foreligger på, er risikoen for ”falsk tryghed” endnu større (se fx Johnson, 1998). Udover tidligere eksempler på faldgruber ved tilbageoversættelse – om franske mænds ”seksuelle” eventyr, blinde og demente – hidrører yderligere et eksempel fra Harkness (i: Harkness et al., 2003). *Das Leben in vollen Zügen geniessen* (Likert-skala) oversættes til *Enjoy life in full trains*, der tilbageoversættes til *Das Leben in vollen Zügen geniessen*. Derpå konkluderes, at oversættelsen er

---

35. ”Misgivings about the usefulness of back-translation on the part of survey research are not new” (Alwin et al., 2002:297).

tilfredsstillende, hvilket er fejlagtigt, eftersom den tyske formulering betyder: *Live life to the full*.<sup>36</sup>

Det afgørende problem er, at tilbageoversættelse ikke – eller i al fald kun i begrænset omfang – tager kulturelle og kontekstuelle forholds betydning ad notam: "... back translation is a method of achieving linguistic equivalence and does not take account of more contextual factors" (Heath et al., 2005:321; se fx også Harkness & Schoua-Glusberg, 1998; O'Shea et al., 2005). Hertil kommer, at spørgsmål ofte kommer til at virke "unaturlige": "... the procedure puts a premium on literal translations, which can lead to unnatural wording of the questions in the target language. Usually the procedure pays more attention to the semantics and less to connotations, naturalness, and comprehensibility" (Rode, 2005:17; se også Van der Vijver & Leung, 1997). Når tilbageoversættelse vælges sigtende imod at opnå "tæt" oversættelse, er den derfor et symptom på særlige opfattelser af, hvad sprog er, og hvordan sprog fungerer blandt sprogbrugere. Mange forfattere efterlyser derfor mere anvendelige procedurer end tilbageoversættelse.

En procedural kontrast til tilbageoversættelse er decentrering, der lægger afstand til "tæt" oversættelse, og hvorom litteraturen ikke er af samme omfang som med hensyn til tilbageoversættelse. Decentrering er en procedure, "... in which the source and the target language versions are equally important and (equally) open to modifications during the translation process" (Brislin (1973), i: Behling & Law, 2000:28f). Med en så bred definition, der ligestiller relevante sprog, muliggør decentrerende oversættelse flere procedurer, fx den af Harkness og Schoua-Glusberg (1998) beskrevne: "Decentering in translation (...) is a technique which

---

36. Et sidste eksempel hidrører fra Usunier (1999): "A good example of how back-translation can be used as an investigation instrument rather than a mere technical device is given by Levine's (1988) comments about research into Brazilian versus US concepts of punctuality. He takes the example of the translation into Portuguese of a questionnaire containing the verb "to wait": "Several of our questions were concerned with how long the respondent would wait for someone to arrive versus when they hoped the person would arrive versus when they actually expected the person would come. Unfortunately for us, it turns out that the term to wait, to hope and to expect are all typically translated as the single verb *esperar* in Portuguese. In many ways our translation difficulties taught us more about Brazilian-Anglo differences in time conception than did the subjects' answers to the questions" (Levine, 1988, pp. 48-9). There is a sort of continuum across languages in the accuracy of description of the waiting phenomenon. The French language, which lies somewhere between English and Portuguese in terms of temporal accuracy, uses two words: *attendre* (to wait) and *espérer* (to hope). To expect has no direct equivalent in French and must be translated by a lengthy circumlocution (*compter sur l'arrivée de*)".



begins from a draft questionnaire in the source language in order to produce final questionnaires in two languages (source and target) through a process of paraphrase and translation between source language and target language” (p. 98).<sup>37</sup> Ifølge sidstnævnte forudsætter decentrering simultant spørgeskemakonstruktion over grænser eller på tværs af kulturer, men proceduren kan også bruges modificeret, når et masterspørgeskema foreligger (se fx Van de Vijver & Leung, 1997).

Om decentrering, der er forholdsvis udbredt i multisurveypraksis, anføres både fordele og ulemper. Blandt fordelene nævnes fx: ”Decentering permits changes to be made to the original-language version where words or concepts are untranslatable or where words or terms have a fairly restricted meaning in one language but a deeper or secondary meaning in the other” (Weidmer et al., 1999:91; se fx også Smith & Bond, 1998; Rode, 2005).<sup>38</sup> Det drejer sig ofte om abstrakte begreber, fx frihed. Decentrerings fordele synes dog kun opnåelige, hvis et samarbejde mellem aktører med sproglig og kulturel ekspertise etableres, fx teams med undersøgelsesrelevante kvalifikationer (se fx Van de Vijver & Leung, 1997). Til de kritiske røster hører Alwin et al.’s (1994): ”Little consideration is given to the fact that a questionnaire is also a text destined for discourse (...). Thus the fields of text analysis and discourse analysis and cross-cultural differences in these spheres must be taken into account” (p. 34).

Yderligere en procedural mulighed er paralleloversættelse, hvor to uafhængigt udarbejdede oversættelser af samme skema sammenlignes for at nå frem til en ”optimal” version (se fx Behling & Law, 2000; Bullinger, 2003; Harkness & Schoua-Glusberg, 1998). Hvis der er væsentlige forskelle mellem de to versioner, træffer forskere afgørelse om den bedste version i samarbejde med oversættere. Nogle forfattere foretrækker proceduren frem for tilbageoversættelse: ”Compared to back translation, there is more emphasis on optimizing wording in the target language” (Smith, i: Presser et al. 2004a:448). Det nævnes også, at paralleloversættelse kan gennemføres ekspedit, da to oversættelser kan laves samtidig. Andre mener, at parallel- og tilbageoversættelse med fordel kan kombi-

---

37. Harkness og Schoua-Glusberg (1998) anskuer parafraseringer som vigtige til sikring af, at spørgeskemaer ikke kun forankres i én kultur.

38. ”Some ‘decentering’ of the concepts in the measure may be necessary, since a literal translation may not capture differences in the nuances of meaning that are found in different cultures” (Smith & Bond, 1998:33).

neres: "Parallel translation combined with back-translation may ... let emic meaning emerge because of different insights among independent translators" (Usunier, 1999).

Et sidste procedureeksempel er komitéoversættelse, som adskillige forfattere anbefaler. Hvad de egentlig anbefaler, er dog ikke éntydigt, eftersom proceduren ikke beskrives på samme måde i litteraturen. Ifølge Smith (2004) er komitéoversættelse enbestydende med, at "... a team of translators and researchers discusses the meaning of items in the source language, possible translations in the target language, and the adequacy of the translations in the target language relating to such matters as level of complexity and naturalness, as well as meaning" (p. 9). Til arbejdsopgaverne hører derfor ikke kun at nå frem til "gode" oversættelser, men "gode" spørgsmål og svarkategorier (se også Smith, i: Presser et al., 2004a).

I anden sammenhæng synes komitéoversættelse imidlertid at være mindre bredspektret og beslægtet med paralleloversættelse: "Committee or parallel translation involves several translators who make independent translations of the same questionnaire (...). At a reconciliation (...) meeting, translators and a translation coordinator compare the translations, reconcile discrepancies and agree on a final version" (Harkness & Schoua-Glusberg, 1998:111).<sup>39</sup> Det er heller ikke éntydigt, hvorvidt tilbageoversættelse kan være et komitéproceduralt aspekt (se fx Van de Vijver & Leung, 1997).

Trods forskellige opfattelser af, hvad komité- eller teamoversættelse er, bliver tilbage adskillige anbefalinger af teamarbejde – både i almindelighed og i særdeleshed ved oversættelse af spørgeskemaer (se fx Behling & Law, 2000; Harkness et al., 2003; Rode, 2005; Smith, 2004; Usunier, 1999). Der er dog også kritiske røster, som fx angår teamsammensætning og -samarbejde. Hvis samarbejdet ikke er tilstrækkeligt åbent og sagligt, er det muligt, at relevante kvalitetssikrende vurderinger glimrer ved deres fravær (se fx Harpaz, 2000).

Uanset valg af oversættelsesprocedure – som eksemplificeret ovenfor – bør oversættelse af spørgeskemaer efter mange forfatteres opfattelse

---

39. Yderligere en definition af komitéoversættelse er: "... two or more translators are used for each translation instead of just one. In addition, a committee including the researchers, the translators, and one or more bilingual reviewers conducts the review of the instrument" (Weidmer et al., 1999:91). Den afsluttende bestræbelse er at opnå oversættelsesmæssig konsensus.

afprøves, før de anvendes i multisurveysk praksis, fx ved gennemførelse af prøveinterview eller kognitive laboratorieinterview. Da dette er en del af en mere omfattende problematik, nemlig afprøvning af spørgeskemaer i almindelighed, henvises behandlingen heraf som nævnt til kapitel 5.



# RESPONDENTERS OPGAVELØSNINGER

## 4.1. INDLEDNING

Rapportens tredje problemstilling handler om respondenters semantiske og andre kognitive opgaveløsninger, der er afgørende for svar, som respondenter afgiver på tværs af grænser eller kulturer:

- Hvordan tages der højde for respondenters opgaveløsninger til sikring af anvendelige data over grænser og på tværs af kulturer?

Som tidligere sættes der fokus på forskelle mellem multi- og monosurveysk litteratur. Det drejer sig om litteratur, der overskrider erfaringsbaseret common sense og forsøger at basere spørgeskemakonstruktion og kvalitetssikring af surveydata på empiriske forskningsresultater.<sup>40</sup>

I store dele af den tidligere kondenserede monosurveyske metodelitteratur fordres, at tilvirkning af surveydata standardiseres (se afsnit 3.2.). Heraf følger, at ”gode” spørgeskemaer er ”... applicable to every respondent or situation that might be encountered” (Aldridge & Levine, 2001:27). Det indebærer tillige, at spørgsmål ”... should be answered in

---

40. Samme værk eller artikel kan kombinere erfarings- og forskningsbaseret, hvorfor samme titler kan forekomme i dette og forrige kapitel.

the same way on different occasions if given to the same person (assuming that the person has not changed in the meantime)” (de Vaus, 2002:96; se også Fink, 1995d; Sapsford, 1999). Andre forskere – fx forskere med tilknytning til CASM-bevægelsen (*Cognitive Aspects of Survey Methodology*) – fremhæver, at standardiseret tilvirkning af surveydata er virkelighedsfjern, endog naiv (se fx Biemer et al., 1991; Braverman & Slater, 1996; Groves, 1989; Lyberg et al., 1997; Schaeffer & Presser, 2003; Schwarz & Sudman, 1996; Sirken et al., 1999; Sudman et al., 1996; Tourangeau et al., 2000).

Såvel i mono- som i multisurveys er tilvirkning af data et komplekst forehavende, der ikke mindst skyldes respondenters vekslende opgaveløsninger: ”Respondents use whatever they perceive as pertinent to come to conclusions about their expected role in the survey interaction and how they should answer, including information not considered by the researcher or questionnaire designer” (Braun, i: Harkness et al., 2003:59). Det er derfor nødvendigt at være forbeholden over for ”... uncritical acceptance of observed differences in the social domain as reflecting valid cross-cultural differences” (Van de Vijver & Leung, 2000:35). Ét potentielt bidrag til måleproblemer er, at tilvirkning af data baseres på respondenters løsning af semantiske opgaver, dvs. forståelse af spørgsmål og svarkategorier. At respondenter i almindelighed skulle forstå spørgsmål på samme måde, er urealistisk selv i monosurveys, hvorfor besvarelse af ”samme” spørgsmål utilsigtet kan generere målinger af forskellige sociale fænomener. Denne problematik behandles i det følgende afsnit, hvor der først sættes fokus på monosurveyforskningslitteratur og dernæst på multisurveyforskning (afsnit 4.2.).

Yderligere en kognitiv opgave er respondenters genkaldelse af ønskede informationer, hvilken fx kan hæmmes af simpel forglemmelse, kognitiv mangel eller kulturelt specifikke genkaldelsesstrategier (afsnit 4.3.). Hertil kommer, at også spørgsmåls vekslende sammenhæng – fx deres succession – kan påvirke afgivne svar i form af konteksteffekter (afsnit 4.4.). Endelig er det ingenlunde sikkert, at respondenter uden videre formidler selv uproblematisk genkaldte informationer, men redigerer dem, før svar afgives, fx under hensyn til, hvad der menes ”socialt ønsket” eller ved simpelthen at afgive usandfærdige informationer (afsnit 4.5.). Forskelle og ligheder mellem mono- og multisurveyforskning sammenfattes og problematiseres i rapportens afsluttende kapitel (afsnit 6.2.).

## 4.2. SPØRGSMÅLSFORSTÅELSE

Afsnittets indledende problematik er monosurveysk metodeforskning, der fokuserer på spørgsmålsforståelse som årsag til måleproblemer: Hvordan forstår respondenter surveyspørgsmåls enkelte ord? Hvordan forstår respondenter hele spørgsmål? Hvordan forstår og tilpasser respondenter deres svar til svarkategorier? (se *oversigt 4.1.*) Som fremstillingen viser, lægges der hermed afstand til anbefalinger som: ”... we want questions that are meaningful, clear, unambiguous, sensitive and revealing” (Aldridge & Levine, 2001:98). Dertil er spørgsmålsforståelse for kompleks. Efter at have besvaret ovennævnte tre spørgsmål rettes opmærksomheden imod tilsvarende multisurveysk metodeforskning, der dog er mere sporadisk.

### Oversigt 4.1.

Aspekter af spørgsmålsforståelse ifølge monosurveysk metodeforskning.

Enkeltord	Spørgsmålshelheder	Svarkategorier
Ords betydningsvidde	Grammatisk struktur	Åbne spørgsmål
Kvantificerende adverbier	Spørgsmålssemantik	Betydningskabelse
Ord og spørgsmålstyper	Spørgsmåls længde	Ved ikke-kategorier
Virkninger af ordforståelse		På midten-kategorier
		Registereffekter

### Monosurveysk metodeforskning

Mange monosurveyske forskere peger på spørgsmålsforståelse som én vigtig årsag til måleproblemer. Eksempelvis pointerer Bradburn & Sudman, at ”... question wording effects can be the greatest single source of bias in surveys” (Bradburn & Sudman, 1989:146; se fx også Fowler, i: Bickman & Rog, 1998). Det af Bradburn & Sudman påpegede problem har været kendt i over et halvt århundrede: ”Since the beginning of scientific surveys, researchers have been sensitive to the fact that changes in the wording of questions could produce dramatic changes in the distribution of responses” (Bradburn, 2000).

Monosurveysk metodeforskning, der underbygger, at *valg af ord* påvirker forståelse af spørgsmål og dermed kan skabe måleproblemer, er omfattende (se fx Alwin, 1991; Biemer et al., 1991; Bradburn, 2000; Fowler, 1993; Schwarz et al., 1991; Smith, 1987). Selv små ændringer i ordvalget kan ændre forståelse af spørgsmål med signifikant forskellige

svarfordelinger til følge. Begrænsning af ords betydningsvidde kan derfor kræve yderligere ord, hvilket kan føre til andre måleproblemer.

Måleproblemer, der skyldes ordforståelse, er både påvist ved besvarelse af faktuelle spørgsmål og holdningsspørgsmål, men oftest ved måling af holdninger. Det synes ikke afgørende, om data tilvirkes ved besøg, telefon eller postvæsen (se fx Schwarz et al., 1991; Wilson et al., i: Schwarz & Sudman, 1996). Med hensyn til faktuelle spørgsmål har Fowler (1992) fx påvist, at måleproblemer ikke kun opstår ved brug af ”vanskelige” og sjældne ord.

Ved besvarelse af faktuelle spørgsmål kan brug af kvantificerende adverbier føre til særlige måleproblemer, fordi den slags ord almindeligvis forstås forskelligt af respondenter (se fx Bradburn & Miles, 1979). Flere metodeforskere har afdækket respondenteres forståelse af kvantificerende adverbier, som Bradburn & Sudman kalder *vage kvantifikatorer*. Ved efterfølgende at operationalisere *ofte* – fx ved at spørge: "Hvor ofte er ofte?" – påpeges betydelige forskelle i forståelse af denne kvantifikator (se også Olsen, 2001a).

De fleste eksempler på måleproblemer, der skyldes ordforståelse, angår holdningsspørgsmål. Ét ofte citeret *split sample*-eksperiment er svarforskelle som følge af brug af verbet (ikke) at *tillade* alternativt at *forbyde* noget (se fx Bradburn, 2000). Siden eksperimentet blev gennemført første gang, er det gentaget flere gange og altid med samme udfald (se fx Schuman & Presser, 1996). Amerikanske respondenter vil hellere *ikke* tillade end forbyde noget. Det er dog ikke altid, at valg af ord påvirker afgivne svar. Eksempelvis har Schuman & Presser (1996) gennemført et *split sample*-eksperiment, hvor nogle respondenter blev anmodet om at artikulere holdninger til abort og andre til graviditetsafbrydelse, og som ikke førte til signifikante svarforskelle. Det sidste eksempel viser, at valget af semantisk analoge ord kun nogle gange, men ikke altid fører til måleproblemer. Hvor ofte og under hvilke omstændigheder, respondenteres forståelse af enkeltord bidrager til at generere måleproblemer, synes fortsat uafklaret, hvorfor udvikling af metoder, der kan bidrage til afklaring af, hvornår den slags måleproblemer er forventelige, efterlyses (se fx Schuman & Presser, 1996).

Da ord har betydningsvidde, gælder tilsvarende *spørgsmålsbelheder*. I praksis er sondringen mellem ord- og spørgsmålsforståelse en abstraktion. Blandt andet derfor retter amerikanske og andre monosurveyske metodeforskere samtidig opmærksomheden mod spørgsmålsforståelse.



Anskuet på tværs peger forskningen på, at måder, hvorpå respondenter forstår spørgsmål, skaber måleproblemer (se fx Schuman & Presser, 1996). I enkelte undersøgelser sættes der fokus på grammatik, men størstedelen angår semantik. Endelig er nogle forskere optaget af, om og hvordan spørgsmålslængde påvirker måleproblemers art og omfang (se *oversigt 4.1.*).

Hvad spørgsmåls *grammatiske struktur* angår, peger flere forfattere på, at strukturen bør være sproglig og ekskludere dobbelte negationer. Med hensyn til sammenhænge mellem grammatik og spørgsmålsforståelse er det velkendt, at syntaktiske strukturer kan ændre spørgsmåls betydning. Imidlertid har monosurveysk forskning kun bidraget marginalt til belysning af problematikken (se fx Tourangeau et al., 2000). Enkelte eksperimenter peger dog på sammenfletninger af syntaks og semantik.

Når det handler om *spørgsmålssemantik*, er monosurveysk metodeforskning omfattende. Det fremgår, at forståelse både af faktuelle spørgsmål og holdningsspørgsmål genererer måleproblemer. Men ligesom forståelse af ord især fremkalder måleproblemer ved besvarelse af holdningsspørgsmål, gælder tilsvarende respondenteres forståelse af spørgsmål (se fx Fowler, 1992; Groves, 1989; Rasinski, 1989; Schuman & Presser, 1996; Schwarz & Sudman, 1996; Sudman & Bradburn, 1982). Adskillige undersøgelser peger på, at moderate ændringer af holdningsspørgsmål fremkalder signifikant forskellige svar og svarfordelinger.

Siden 1970'erne har især amerikanske forskere gennemført mange undersøgelser af måleproblemer, der skyldes spørgsmålsforståelse (se fx Bradburn & Sudman, 1989; Schuman & Presser, 1977; Schuman & Presser, 1981; Sudman & Bradburn, 1987; Sudman & Bradburn, 1982). Under ét bidrag 1970'ernes og 1980'ernes ofte *split sample*-baserede forskning til problematisering af respondenter som medier for tilvirkning af surveydata og førte samtidig til ledetråde for udarbejdelse af spørgsmål. Converse & Presser anbefaler fx "simple language and common concepts" og efterlyser "simple questions rather than general ones" til lige med "clear definitions for terms of shared meaning" (Converse & Presser, 1986).

Trods amerikansk dominans er udforskning af spørgsmålsforståelse langt fra et rent amerikansk anliggende. Eksempelvis afprøvede en engelsk forsker omkring 30 spørgsmål ved intensive interview med 265 respondenter (Belson, 1981). Da næsten alle spørgsmål gav anled-

ning til problemer, konkluderede Belson, at en væsentlig årsag til måleproblemer er, at respondenter ofte ikke forstår spørgsmål og derfor ikke er i stand til at afgive "sande" svar. Konklusionen førte til formulering af 15 hypoteser, fx: Spørgsmål, der er vanskelige at forstå for respondenter, forstås på måder, der gør dem lettere at besvare, ligesom spørgsmålsforståelse tilpasses respondenteres egne forhold. Hertil kommer en tredje hypotese om, at dele af spørgsmål under visse omstændigheder ignoreres.

Siden 1990'erne har amerikanske og andre forskere ydet fortsatte bidrag til udforskning af, hvordan respondenter forstår spørgsmål, og hvordan og i hvilket omfang forståelse skaber måleproblemer (se fx Lyberg et al., 1997; Schwarz & Sudman, 1996; Sirken et al., 1999; Sudman et al., 1996; Tanur, 1992; Tourangeau et al., 2000). Sudman et al. (1996) påpeger fx, at spørgsmålsforståelse både fordrer fokus på *literal meaning* og *intended meaning*. "Even if comprehending the literal meaning of the question poses no problem, respondents still have to infer what the specific information is that the researchers are interested in" (Sudman et al., 1996:63f). I så henseende gør respondenter fx brug af svarkategorier, når disse kendes, ligesom også konteksten spiller en rolle. Derfor betyder "ens" spørgsmål ikke det samme for respondenter, men varierer fx med hensyn til social og kulturel baggrund (tilsvarende, se fx Johnson et al., i: Lyberg et al., 1997).

I et andet monosurvey-skift påpeges ligeledes surveydatas skrøbelighed: "As survey researchers have known for some time, many reporting problems arise because respondents misunderstand the questions" (Tourangeau et al., 2000:9). Ifølge forfatterne er spørgsmål altid kendetegnet ved *a space of uncertainty*. Forfatterne sondrer mellem to opfattelser af spørgsmålsforståelse, hvoraf den sidste bifaldes. Den første er *representation of the sentence*, mens den anden er *representation about the sentence*. Som hos Sudman et al. (1996) er spørgsmålsforståelse ikke *representation of the sentence*, da respondenter drager slutninger angående *representations about the sentence*, hvilket får afgørende betydning for respondenteres svar og genererer måleproblemer: "There is an endless set of possible inferences that respondents can make about the question ..." (Tourangeau et al., 2000:33).

Ifølge Tourangeau et al. (2000) er opgaven at reducere det "semantiske rum", hvilket fordrer, at spørgsmål er så entydige som muligt, fx ved fravalg af kompleks syntaks, diffuse ord og begreber, vage adver-

bielle kvantifikatorer og utilsigtede forudsætninger. Hvis vage eller sjældne ord er uundgåelige, bør de defineres eller eksemplificeres. Men uanset bestræbelser i retning af ”entydighed” kan spørgsmål næsten altid forstås på flere måder: “... unambiguous questions hardly exist, due to the intrinsic ambiguity of language” (Houtkoop-Steenstra, 2000:8),

Et sidste aspekt af spørgsmålsforståelse er *spørgsmåls længde*. Mange anbefaler som nævnt spørgsmål med få ord (se afsnit 3.2.). Men er anbefalingen evident? Monosurveysk metodeforskning tilbyder ikke et entydigt svar. Lange spørgsmål kan fremkalde mere valide målinger end korte, fx ved besvarelse af spørgsmål om sygdom og rusmiddelvaner (se fx Bradburn & Sudman, i: Biemer et al., 1991; Groves, 1989). Bateson (1984) påpeger, at lange spørgsmål, der fx præciserer vage ord og termer, kan reducere antallet af betydningsudfald. Andre eksperimenter fører til anbefaling af vekslen mellem lange og korte spørgsmål. Endelig peger et eksperiment på, at mellemlange spørgsmål med omkring 20 ord er at foretrække.

Det er for så vidt ikke overraskende, hvis lange spørgsmål kan være at foretrække. Det skyldes, at betydningsudfald kan reduceres ved at øget spørgsmåls længden (se fx Olsen, 2001a). Måske er lange spørgsmål tillige bedre end korte til at fremme genkaldelse af informationer? Hermed peges på problemets kompleksitet. Antallet af ord beror ikke kun på overvejelser om spørgsmålsforståelse, men også på muligheder for genkaldelse af informationer. Ifølge flere monosurveyske metodeforskere fremmer lange spørgsmål genkaldelse og kan samtidig øge respondenters svarmotivation (se fx Crano & Brewer, 2002; Fowler, i: Bickman & Rog, 1998).

I surveys er spørgsmål normalt lukkede, dvs. efterfølges af *svarkategorier*. De fleste monosurveyske metodeforskere synes da også enige om, at *åbne spørgsmål* bør anvendes moderat. Én årsag er surveys standardiserende ”logik”, en anden pragmatisk: “Data from open-ended questions are much more time consuming to read, reduce and analyze” (Nolinske, 1998). Det kan også spille en rolle, at adskillige undersøgelser viser, at ”samme” åbne og lukkede spørgsmål ikke fører til statistisk kongruente svarfordelinger (se fx Schwarz & Hippler, i: Biemer et al., 1991).

Der er ikke konsensus om, hvornår åbne spørgsmål bør anvendes. I nogle undersøgelser vises, at ”socialt uønsket” adfærd kan belyses ved brug af åbne spørgsmål (se fx Bradburn & Sudman, 1979). Undertiden kan også holdningsmålinger forbedres ved brug af åbne spørgsmål

(se fx Geer, 1991). På den anden side pointerer fx Hippler et al. (1987), at åbne spørgsmål kan skabe måleproblemer, fordi det selvindlysende fravælges. Kritikken af lukkede spørgsmål er ikke forstummet: "Recent research has shown that there are distinct disadvantages to closed-ended questions, and open-ended questions are not as problematic as they seemed" (Krosnick, 1999).

Hvilken rolle spiller svarkategorier for respondenters opgaveløsninger, når spørgsmål er lukkede? Det afhænger af, om respondenter informeres om kategorierne. Når kategorier er ukendte, er spørgsmål "åbne" anskuet fra respondenters synsvinkel. Når respondenter kender svarkategorier, udgør de en del af den samlede forståelsesramme. Antallet af ord, som skal fastholdes i arbejdshukommelsen, øges, men samtidig muliggøres reduktion af mulige betydningsudfald (se fx Olsen, 2001a). To "identiske" spørgsmål, hvoraf det ene er åbent og det andet lukket, er derfor semantisk forskellige (se fx Bradburn & Sudman, i: Biemer et al., 1991).

Ligesom forståelse af spørgsmål varierer, gælder det også svarkategorier (se fx Johnson et al., i: Lyberg et al., 1997). At svarkategorier påvirker svar har været kendt længe, hvorimod ikke mange har gjort forsøg på at belyse bagvedliggende kognitive processer. I forlængelse af ovennævnte understreger også Schwarz og Hippler, at svarkategorier ikke er "tekniske", men har stor betydning for måleproblemers omfang: "... response alternatives not only serve to record respondents' answers, but also define respondents' substantive tasks. They therefore influence respondents' interpretation of the questions ..." (Schwarz & Hippler, i: Biemer et al., 1991:43).

Det er derfor afgørende, at også svarkategorier overvejes nøje: "Even when respondents have a clear answer to report, it may not be clear to them how to report it" (Tourangeau et al., 2000:13). Nogle respondenter anstiller fx omhyggelige betragtninger, hvorvidt de skal svare *meget enig* eller *enig*, mens andre er i tvivl om, hvorvidt en given aktivitet er *ofte* eller *meget ofte* forekommende. Andre undlader derimod at reflektere svarkategorielt og træffer ekspeditte valg. Som foreslået af Tourangeau et al. bør problemet løses ved, at svarkategorier bidrager til at reducere det "semantiske rum".

Et særligt svarkategorielt problem er adverbelle kvantifikatorer, som mange forskere advarer imod. Hvis respondenter fx spørges: *Hvor ofte går De i kirke?* og kategorierne er ikke-adverbelle, synes måleproble-

merne at blive reduceret sammenlignet med anvendelse af adverbelle kvantifikatorer, fx *ofte* og *sjældent*. Undersøgelser peger på, at svarkategorier af sidstnævnte type ikke forstås på samme måde (se fx Bradburn & Miles, 1979; Bradburn & Sudman, 1979; Olsen, 2001a; Singleton et al., 1988).

Når det handler om *ved ikke*-svarkategorier, mener de fleste metodeforskere, at de ikke bør anvendes i tilknytning til faktuelle spørgsmål (se fx Groves, 1989). Ikke desto mindre peger undersøgelser på, at svar ændres, når muligheden tilbydes (se fx Poe et al., 1988), hvilket antyder, at respondenter "genkalder" informationer, som de ikke har genskabt. Med hensyn til holdningsspørgsmål mener mange, at *ved ikke*-kategorier bør tilbydes, selv om muligheden via "holdningsbortfald" fører til holdningsunderrapportering (se fx Crano & Brewer, 2002; Schwarz & Hippler, i: Biemer et al., 1991). Ikke alle har nemlig holdninger til alt, og for så vidt er en ikke-holdning også en holdning.<sup>41</sup>

Også svarkategoriernes *antal og rækkefølge* kan skabe måleproblemer via *registereffekter*, der er dokumenteret gentagne gange (se fx Bradburn, 2000; Lyberg et al., 1997; Schwarz & Sudman, 1992; Sudman et al., 1996). Fænomenet synes vanskeligt at forklare. Dammen og de Bie (i: Schwarz & Sudman, 1992) peger fx på, at registereffekter kan skyldes et samspil mellem forståelse og vurdering af emnelister, som respondenter tilbydes. I spørgsmål, hvor respondenter skal vurdere forskellige emner, som præsenteres på en liste, kan forståelse af emner, der står øverst, danne en semantisk ramme, som påvirker forståelse af efterfølgende emner. Forståelsesoverførsel kan bevirke, at det mest "tiltalende" emne tildeles det mest positive svar.

Afhængigt af antal svarkategorier, kategoriernes art og af proceduren anvendt ved tilvirkningen af data (dvs. post, besøg eller telefon), forekommer både førsteffekter (*primacy effects*) og sidsteffekter (*recency*

---

41. Angående neutrale *på midten*-svarkategorier brugt i forlængelse af holdningsspørgsmål udviser forskningen ikke éntydige tendenser. Hvis et spørgsmål fx er: *Hvor vigtig mener De, begavelse er, hvis man skal opnå en høj social position i samfundet?*, kan svarkategorierne *meget vigtig*, *noget vigtig* og *ikke særlig vigtig* tilbydes, men en fjerde kategori – *både vigtig og ikke vigtig* – kan tilføjes. Når på midten-kategorier tilføjes, ændres svar, hvilket er påvist i flere *split sample*-eksperimenter (se fx Olsen, 2001a; Schuman & Presser, 1996, og Schwarz et al., 1991). Eksempelvis gennemførte Schuman & Presser (op.cit.) eksperimenter, hvor hver anden respondent fik tilbudt *på midten*-kategorier. Respondenterne i de to subsampler udviste signifikant forskellig svaradfærd. Ikke blot vælger flere respondenter på midten-svar, når muligheden foreslås, men samtidig fremkommer færre pol-svar.

*effects*) (se fx Schuman & Presser, 1996): "... response order effects are assumed to depend on the item's serial position, their plausability, and the administration mode used" (Schwarz et al., 1991). Undersøgelser viser, at førsteffekter forekommer, når respondenter får udleveret svarlister (se fx Bradburn, 2000). Førsteffekter beror på, at respondenter vælger initialt præsenterede emner, som forekommer rimelige. Hvis alle emner synes rimelige, foretrækker respondenter ofte emner øverst på listerne (se fx Krosnick, i: Schwarz & Sudman, 1992). Sidsteffekter forekommer derimod, når svarkategorier læses op, fx telefonisk (se fx Bradburn, 2000).

### **Multisurveysk metodeforskning**

Efter at have præsenteret monosurveyske forskningsresultater om respondenters forståelse af enkeltord, hele spørgsmål og svarkategorier sættes der fokus på variationer i spørgsmålsforståelse over grænser og på tværs af kulturer. I så henseende er metodeforskningen af mindre omfang, hvorfor de følgende eksempler bliver tilsvarende færre. Adskillige forskere peger på, at variationer i "grænseoverskridende" forståelse af ord, spørgsmål og svarkategorier er vigtige årsager til ækvivalensproblemer: "Vocabulary (...), semantics, and pragmatic meaning of words and utterances do not match up neatly across languages" (Harkness et al., i: Presser et al., 2004a:456; vedr. forståelse af spørgsmål og svarkategorier, se fx Blais & Gidengil, 1993; Braun & Scott, 1998; Cheung & Rensvold, 2000; Clarke, 2001; Harzing, 2005; Harzing, 2006; Johnson et al., 1997; Mohler et al., 1998; Schaffer & Riordan, 2003; Smith, 2004; Smith, 2004b; Smith, i: Harkness et al., 2003; Usunier, 1999; Van Herk et al., 2004). Samtidig understreger flere forskere vanskelighederne ved at kunne forudsige effekter af spørgsmålsforståelse.

Nogle af de nævnte forskere underbygger empirisk, hvordan vekslende forståelse af ord og spørgsmål genererer ækvivalensproblemer, men opmærksomheden rettes også imod, hvordan forståelse af svarkategorier påvirker svarprocessen og skaber usammenlignelige data (se *oversigt 4.2*). I det følgende nævnes enkelte eksempler på videnskabelige artikler og andre fremstillinger, som viser, at forståelsesbetingede måleproblemer, der i høj grad er sandsynliggjort i monosurveysk metodeforskning, ikke overraskende tillige gør sig gældende, når grænser og/eller kulturer overskrides. Der foreligger således næppe en fundamentalt anderledes

metodeproblematik, men muligvis endnu større måleproblemer end i monosurveys.

## Oversigt 4.2.

Aspekter af spørgsmålsforståelse ifølge multisurveyssk metodeforskning.

Spørgsmål	Svarkategorier
Ords betydningsvidde	Kategoriens betydningsvidde
Spørgsmåls betydningsvidde	På midten-kategorier Skalerende kategorier Simple kategorier Ved ikke-kategorier

En amerikansk undersøgelse handler om, hvorvidt spørgsmålsforståelse afhænger af kulturelt tilhørsforhold, dvs. blandt amerikanere med afrikansk, mexicansk, puertoricansk og hvid europæisk baggrund (Johnson et al., 1997). Ved respondent-*probing* undersøges, hvordan forskellige spørgsmål forstås, fx følgende: *During the past 2 weeks, would you say you experienced a lot of stress, a moderate amount of stress, relatively little stress, or almost no stress at all?* Når cirka 500 respondenter derpå spørges: *In this question, what does the word 'stress' mean to you?*, er forståelserne kulturelt forskellige, fx med hensyn til om sociale relationer eller økonomiske problemer inkluderes i stressbegrebet eller ikke. Tilsvarende kulturspecifikke forskelle viser sig også ved respondent-*probing* i tilknytning til flere andre spørgsmål, fx: *How much control do you think you have over your future health?* Konsekvensen af, at respondenteres kulturelle tilhørsforhold er afgørende for spørgsmålsforståelse, bliver, at utilsigtede socio-kulturelle forskelle måles – ikke faktiske fænomener (op.cit.).<sup>42</sup>

Alt i alt konkluderer Johnson et al. (1997), at "... when more than two cultures participate in a survey, the array of culturally mediated contextual elements that underlie question interpretations becomes multidimensional. What is found to be emic and what is shared cross-culturally appear to vary depending on which cultures and which concepts are being investigated" (p. 106). Kulturelle forskelle i eller mellem lande, der deltager i multisurveys, genererer derfor yderligere ækvivalens-

---

42. En mindre undersøgelse fra Chicago-området underbygger ligeledes, at respondenteres etniske tilhørsforhold – afrikansk, mexicansk og europæisk baggrund – påvirker spørgsmålsforståelse, fx med hensyn til "fysiske aktiviteter" (Warnecke et al., 1997).

problemer, hvilket ikke er overraskende for Johnson et al. (op.cit.). De peger bl.a. på, at kulturelle opfattelser længe har vist sig at påvirke betydninger, som mennesker tildeler tekster. Det er fx velkendt, at kulturelle grupper er uens med hensyn til grad af kollektivism versus individualisme og grad af følelsesmæssig kontrol og ekspressivitet, ligesom der også er forskelle, hvad maskulinitet versus femininitet angår. Hertil kommer, at også abstraktionsevne er kulturelt betinget (se fx Scheuch, 1993). Ikke desto mindre fortsætter megen surveyforskning med at ignorere problemet vedrørende tværkulturel ækvivalens, dvs. sonderer ikke klart mellem universelle ”etiske” og kulturspecifikke ”emiske” begreber: ”The widespread use of pseudoetic constructs has been an important criticism of survey research for many years (Johnson et al., 1997:89).

En canadisk undersøgelse omhandler ækvivalensproblemer mellem en engelsk- og fransksproget version af ”samme” spørgeskema, hvis versioner blev udarbejdet parallelt og tilsyneladende prætestet omhyggeligt (Blais & Gidengil, 1993). Forskerne var derfor overbeviste om, at undersøgelsen ville være uden forståelsesbetingede ækvivalensproblemer, men det modsatte gjorde sig gældende. Selv forståelse af almindelige hverdagsord som *things* i forhold til *la situation*, udløste signifikant forskellige svarfordelinger, som Blais og Gidengil forklarer med henvisning til ordenes konnotationer, hvorpå de som andre multisurveyske metodeforskere konkluderer: ”Clearly, linguistic equivalence is difficult to achieve” (1993:553).

Det kunne også (fejlagtigt) formodes, at et udbredt ord som *Gud* ville blive forstået på nogenlunde samme måde på tværs af grænser, hvilket langtfra er tilfældet. Når fx amerikanere og japanere spørges: *Do you believe in God?*, svarer næsten alle amerikanere (90 pct.) bekræftende, mens et mindretal af japanere afgiver tilsvarende svar (35 pct.) (Sasaki, 1995). Heraf kan dog ikke udledes, at de fleste japanere er ateister: ”... probing reveals that many Japanese who answer ‘no’ nonetheless act in ways which Westerners would otherwise consider religious. This is but one myriad instances in which differing social structural and cultural contexts significantly impact the results of cross-national attitudinal surveys” (op.cit.:105). Da det (næsten) er umuligt at oversætte *Gud* til japansk, forklares den japanske ”ateisme” bl.a. ved problemer med at ”transportere” ord fra ét sprog til et andet (se fx Heath et al., 2005; Jowell, 1998).

Et undersøgelseseksempel, der især omhandler ex post-kontrol af funktionel ækvivalens, sætter fokus på manglende sammenlignelighed



bl.a. mellem engelske, italienske og tyske ISSP-data (Braun & Scott, 1998).<sup>43</sup> Ex post-analyserne, der baseres på data om respondenteres stil-  
lingtagen til en række påstande og oversættelser heraf, fx: *All in all, family  
life suffers when the woman has full-time job* og *A woman and her family will all be  
happier if she goes out to work*, sandsynliggør ækvivalensproblemer, der skyl-  
des problematiske oversættelser af allerede i udgangspunktet tvetydige  
spørgsmål, men tillige kulturelle forskelle: ”Perhaps the most important  
aspect ... is the way social reality and language interact. Identical ques-  
tions might be understood differently against different back-  
grounds”(Braun & Scott, 1998:143). Det spørgsmål, som artiklens titel  
rejser (*Is having a job the same as working?*), besvares derfor benægtende.

Yderligere et eksempel på, hvordan vekslende forståelse af over-  
satte spørgsmål genererer ækvivalensproblemer, angår et ISSP-spørgsmål  
om oplevelse af lykke: *If you were to consider your life these days, how happy or  
unhappy would you say you are, on the whole ... very happy, fairly happy, not very  
happy, not at all happy, can't choose?* (Smith, 2004:5; se også Smith, 2004b).  
Smith peger på, at spørgsmålet sandsynligvis ikke har samme betydning  
på andre europæiske sprog, hvor det fx kan konnotere til tilfredshed.  
Heller ikke *very, fairly* og *not at all* har samme betydning, når ordene over-  
sættes til andre europæiske sprog. Det er fx tvivlsomt, om *very* og det  
franske *très* er ensbetydende. Med henvisning bl.a. til Grunert & Muller  
(1996) peger Smith på et behov for præcise empiriske indikatorer, når  
ord er abstrakte og derfor – som i tilfældet lykke – har stor betydnings-  
vidde: ”Even given the most careful translation, it is difficult to compare  
distributions of two questions that employ abstract concepts and subjective  
response categories” (Smith, i: Harkness et al., 2003:72).

En finsk undersøgelse peger på, at spørgsmål om alkoholvaner  
ikke forstås på samme måde blandt respondenter i Finland, Tyskland,  
Italien og Holland (Raitasalo et al., 2005). Undersøgelsen, der er baseret  
på kognitive interview, peger på: ”... important differences related to  
cultural understanding on certain specific questions, perhaps due to dif-  
ferent cultural positions of alcohol in the four countries. As for the last  
point, the overall interpretation is that in Germany and Italy, routines  
and practices where drinking may take place are more closely integrated  
in everyday life than in Finland or the Netherlands, and as such, are also

---

43. Ex post-procedurer til kontrol af funktionel ækvivalens kan fx være faktoranalyse, psykometriske  
tilgange eller multidimensionel skalering (MDS).

more easily and in a more unproblematic way available as reference points” (op.cit.).

At sprog, som spørgeskemaer oversættes til, i sig selv kan generere ækvivalensproblemer – her tværnationale – underbygges i en australsk undersøgelse (Harzing, 2005). ”Could the language of the questionnaire influence a person’s response?”, spørger Harzing (p. 214) og besvarer spørgsmålet bekræftende. Når universitetsstuderendes tværnationale kulturelle værdier afdækkes og sammenlignes, bliver værdiforskellene mellem 24 lande signifikant mindre, når det anvendte spørgeskemasprog er engelsk. Resultatet er med andre ord en tendens til ”homogenisering”: ”... our study has shown that decision on the language of the questionnaire should be a key aspect of any cross-national study design. Where questions comprise an element of culture – and we have shown that this might be the case even with questions that at first glance would be considered neutral – the use of English-language questionnaires might obscure important differences between countries” (op.cit.:222).

Multisurveyske metodeforskere sætter som nævnt ikke kun fokus på respondenteres forståelse af spørgsmål, men tillige på forståelse af *svarkategorier* og heraf resulterende ækvivalensproblemer (se fx Johnson et al., 1997; Johnson et al., 2005; Möhler et al., 1998; Smith, i: Presser et al., 2004a; Voss et al., 1996). Ét aspekt er oversatte svarkategoriens vekslende betydningsvidde (se *oversigt 4.2*). Ligesom forståelse af spørgsmål kan generere ækvivalensproblemer, gælder tilsvarende svarkategorier. Som bl.a. påpeget af Schwarz & Hippler (i: Biemer et al., 1991) kan svarkategorier påvirke forståelse af spørgsmål og være med til at bestemme, hvilke informationer respondenter oplever at skulle genkalde.

At forståelse af svarkategorier kan være afgørende for, hvad respondenter svarer, viser sig i en undersøgelse, hvor svarkategorier som fx *good* og *poor* oversættes fra engelsk til mandarin og japansk: ”Our data show that standard translation/back-translation of the standard ‘very good, good, so-so, poor, very poor scale’ (...) from English into Putorghua (Mandarin) Chinese and or Japanese would result in extremely different meanings such that the comparability of results would be questionable” (Voss et al., 1996:44). Undersøgelsen, hvis respondenter *rater* adjektiver ved at tildele dem talstørrelser, tyder derfor på, at simple oversættelser af adjektiver fra et sprog til et andet ikke nødvendigvis skaber

ækvivalente betydninger, hvilket bl.a. skyldes adjektivers kontekstuelle afhængighed.<sup>44</sup>

Et andet eksempel med amerikanske og tyske respondenter angår forståelse af svarkategorierne *agree* og *disagree* oversat fra engelsk til tysk (Mohler et al., 1998). Trods umiddelbar semantisk ækvivalens er de oversatte kategorier ikke sammenlignelige med de oprindelige, hvilket viser sig ved markante svarforskelle, der efter Mohler et al.'s vurdering bl.a. forklares ved forståelsesforskelle: "... the simple base terms such as *agree* – *stimme zu*, *disagree* – *lene ab/stimme nich zu* are rated more extremely by German respondents than their English counterparts are by American respondents. It remains to be seen whether this means the German expressions involve greater intensity of agreement/disagreement, etc., or whether, independent of this, German respondents differ in rating behaviour" (op.cit.:178).

Et andet svarkategorielt aspekt angår forståelse og svaradfærd i relation til neutrale *på midten-* versus ekstreme *pol-kategorier* almindeligvis i tilknytning til holdningsspørgsmål, fx *hverken enig eller uenig* i forhold til *meget uenig*. Det er som tidligere nævnt et aspekt, som også monosurveyforskning sætter fokus på, og hvor tendenserne er tvetydige. I multisurveys om respondenters holdninger og værdier antages ofte, at svarforskelle er umiddelbart *face value*-sammenlignelige. Imidlertid kan forståelse og svaradfærd bl.a. i relation til *på midten-* og *pol-kategorier* bevirke, at data ikke bliver ækvivalente:<sup>45</sup> "If members of one culture have an inherent propensity to respond to survey instruments, using only the extreme response categories of the scale, while members of the other cultures habitually exhibit a restricted response style, comparisons between groups will be distorted" (Clarke, 2001:137; se også Cheung & Rensvold, 2000).

Baseret på data om madlavning, opvask og personlig pleje i Grækenland, Italien, Spanien, Frankrig, Tyskland og UK viser en undersøgelse fx, at *pol-tendenser* – *extreme response styles* – er mere udbredt i lande omkring Middelhavet – især i Grækenland, men muligvis ikke Italien – end i det nordvestlige Europa: "This strongly confirms that ignoring

---

44. "Respondents rated an assortment of common adjectives by comparing the magnitude of the word to a given modulus. The results indicate that the traditional translation/back-translation technique may not provide response intervals that are comparable cross-culturally" (Voss et al., 1996).

45. Det gælder også med hensyn til kognitiv magelighed og "føjelighed" (se afsnit 4.3.).

national differences in response styles may lead to invalid inferences in cross-cultural research” (Van Herk et al., 2004:358). Hvad enten forklaringen er semantisk, kulturel eller begge dele, ”tiltrækkes” visse respondenter åbenbart af pol-kategorier, mens andre gerne holder sig fra dem.

På grundlag af data fra 19 lande er der fx fundet tendenser til pol-svar blandt kaukasere, afroamerikanere og latinamerikanere (se fx Harzing, 2006; Johnson et al., 2005). På den anden side undgår asiater – i særdeleshed japanere – ekstreme svar (se fx Smith, i: Harkness et al., 2004). Smith opfordrer derfor til at ”blødgøre” pol-kategorier, når multi-surveys omfatter respondenter, der søger imod ”midten”, fx *tend to agree* i stedet for *agree*. Surveys behøver end ikke at overskride grænser for at respondenter udviser markant forskellige tendenser til pol-svar. I USA er pol-svar fx mere udbredt blandt afroamerikanske end blandt hvide studerende (Smith, i: op.cit.). På samme måde viser en ældre undersøgelse af amerikanske flåderekrutter markante etniske pol-forskelle (Hui & Triandis, 1989).

Nogle forskere forklarer på midten-tendenser, der undertiden truer multisurveys dataækivalens, ved kollektivistiske personlighedstræk, mens pol-tendenser forklares ved individualisme: ”In Asian cultures ... a tendency to respond neutrally to survey items might be related to cultural norms dictating modesty and cautious responses, whereas in Western cultures, a tendency to use extreme responses might be related to norms that endorse individual expressiveness” (Schaffer & Riordan, 2003:192; se også Johnson et al., 1997; Triandis, 1992).<sup>46</sup> Andre peger på, at også kulturel maskulinitet kan spille en rolle for pol-svar (se fx Johnson et al., 2005).

Da *skalerende svarkategorier* – fx i tilknytning til Likert-skalaer eller andre holdningsskalaer, der ofte anvendes i tværkulturelle surveys – har yder- og midterpunkter, er de et aspekt af pol- og på midten-problematikken, der er møntet på, ”... whether respondents perceive and interpret rating-scale intervals in the same manner (scaling equivalence)” (Schaffer & Riordan, 2003:191). Sådan forholder det sig ikke nødvendigvis, hvilket bl.a. afhænger af uddannelsesniveaue og føjelighed

---

46. Triandis (1995:2) definerer ”individualisme” som en ”... social pattern that consists of loosely linked individuals who view themselves as independent of collectives”, mens ”kollektivisme” har karakter af “... social pattern consisting of closely linked individuals who see themselves as parts of one or more collectives”.

(se afsnit 4.3.). I udviklingslande kan ordinale svarkategorier være mere anvendelige, fordi mange respondenter ikke er tilvænnet numeriske *rating*-skalaer (se fx Patel et al., 2005).

Mens nogle forskere anbefaler non-verbale numeriske svarkategorier, fordi de menes at have ”universelle” træk (se fx Fowler, 1995), er andre mere forbeholdne: ”... numerical scales are not as invariant in meaning and error free as their mathematical nature presupposes” (Smith, i: Presser et al., 2004a:435). I en 5-punkts Likert-skala viser 3 sig fx at betyde *no opinion* blandt erhvervsaktive i USA eller *mild agreement* blandt erhvervsaktive i Sydkorea (Schaffer & Riordan, 2003). En anden undersøgelse, viser, at tyske respondenter ved besvarelse af spørgsmålet *How successful would you say you have been in life?* påvirkes statistisk signifikant af, om tilbudte numeriske svarkategorier befinder sig i intervallet 0-10 eller 0-5 (Schwarz et al., 1991; se også Schwarz & Hippler, 1995). De forskellige svarfordelinger kan fx være forårsaget af ”lykkenumre” eller talstørrelsers succession fra venstre mod højre. Selv i monosurveys kan numeriske skaleringer være problematiske, fordi der er usikkerhed om, hvordan de forstås, fx ved større eller mindre afstand mellem positive og negative numeriske svarkategorier: ”... considerable caution should be exercised when using scales developed in one country or cultural context in other environments” (Douglas & Nijssen, 2002; se også Hoffmeyer-Zlotnik, 2004).

Nogle forskere opfordrer derfor til overvejelse af mere *simple svarkategorier*. Det kan fx være dikotome eller visuelle kategorier fx i form af mere eller mindre glade ansigter (se fx Smith, 2004). Patel et al. (2005) og andre mener, at dikotome svarkategorier er særligt anbefalelsesværdige i udviklingslande. Men heller ikke dikotome kategorier – det gælder fx *agree* versus *disagree* – eller visuelle stimuli er nødvendigvis kulturelt ækvivalente (se fx Smith, i: Presser et al., 2004a). Selv hvis sondringen mellem *agree* og *disagree* skulle være entydig på engelsk, kan den komme til at betyde noget andet, når den oversættes til tysk (se fx Mohler et al., 1998). Hertil kommer, at dikotome målinger fører til mindre målenøjagtighed.

Endelig sætter dele af litteraturen fokus på brug af *ved ikke-svarkategorier*, som de fleste monosurveyske forskere anbefaler i tilknytning til holdningsspørgsmål. Det er dog kun en problematik, der strejfes i multisurveysk litteratur. Eksempelvis nævner Smith (i: Presser et al., 2004a; se også Smith, i: Harkness et al., 2003), at afgivelse af ved ikke-svar både varierer over grænser og på tværs af kulturer. Den slags svar

forekommer fx oftere i lande, hvis befolkninger er uden eller med begrænset uddannelse (se fx også Usunier, 1999). Men hvis forskellene er ”virkelige”, fx at flere i nogle lande har mere robuste og bredspektrede holdninger end i andre, giver anvendelse af ved ikke-kategorier ikke anledning til ækvivalensproblemer, påpeger Smith (op.cit.). Problemet er imidlertid, at det sjældent vides, hvornår holdninger er robuste eller skrøbelige ad hoc-konstruktioner.

#### 4.3. GENKALDELSE AF INFORMATIONER

Forståelse af spørgsmål og svarkategorier er en forudsætning for genkaldelse af informationer: ”There are limits to what people are able to recall. If a question calls for information that most people cannot recall easily, the data will almost certainly suffer” (Fowler, i: Bickman & Rog, 1998:352). Som kognitive processer er genkaldelse rekonstruktion af indkodede informationer og retter sig fx imod: ”... such processes as adopting a retrieval strategy, generating specific retrieval cues to trigger recall, recollecting individual memories, and filling in partial memories through inference” (Tourangeau et al., 2000:9). Som tidligere sættes der først summarisk fokus på monosurveysk metodeforskning, hvorpå opmærksomheden rettes imod multisurveysk, der er endnu mere sporadisk end om spørgsmålsforståelse.

##### **Monosurveysk metodeforskning**

Monosurveyske metodeforskere sætter fokus på forskellige aspekter af genkaldelse (se *oversigt 4.3*). Én konceptuel sondring er mellem episodisk og semantisk hukommelse, hvor sidstnævnte er generel videnshukommelse. Et andet aspekt angår tiden fra indkodning til genkaldelse af informationer, mens et tredje er teleskopering, dvs. fejlagtige tidsmæssige placeringer af informationer. Yderligere et aspekt angår det forhold, at nogle hændelser og oplevelser har mere ”karakteristiske træk” end andre. Endelig rejses også spørgsmål om kognitiv magelighed og fremme af pålidelig genkaldelse.

### Oversigt 4.3.

Aspekter af genkaldelse ifølge monosurveysk metodeforskning.

---

- Fra indkodning til genkaldelse af informationer
  - Teleskopering
  - Begivenheders karakteristiske træk
  - Kognitiv mæglighed
  - Kognitive aflastninger
- 

Langtidshukommelse er både *episodisk* og *semantisk*. I førstnævnte tilfælde genkalder informationer i koordinatsystem med tid og rum som akser (se fx Eysenck, 1993). Semantisk hukommelse overskrider derimod det spatio-temporale og omfatter "... the organised knowledge that we possess about language and about the world" (op.cit.; se også Tanur, 1992). Enkelte forskere har gennemført undersøgelser, som peger på, at respondenter generelle hukommelse undertiden tildeles forrang ved besvarelse af episodiske spørgsmål (se fx Burton & Blair, 1991; Tourangeau et al., 2000). Respondenter, der genkalder hændelsesinformationer episodisk, undlader at informere om alle begivenheder, mens respondenter, hvis svar navnlig beror på ikke-episodisk hukommelse, ofte overvurderer antallet af begivenheder. Ved ikke-episodisk genkaldelse baseres "... frequency estimates on a rate of behavioral occurrence without recalling any specific behavioral episodes" (Blair & Burton, 1987:63). Heraf følger, at respondenter undertiden "rekonstruerer" informationer om hændelser, der aldrig har fundet sted.

Tilsvarende har Bower et al. (1979) gennemført eksperimenter om respondenter rutineaktiviteter, og hvorledes hukommelse herom organiseres og aktiveres. Forsøgene peger på, at nogle respondenter genkalder kognitive drejebøger (*scripts*) i stedet for informationer om faktiske hændelser, og at respondenter ofte "husker" hændelser i overensstemmelse med drejebøger. Effekten er, at det oplevede kontamineres med drejebøgers forskrifter (se fx Fowler, 1993:89). Eksemplerne anskueliggør, at spørgsmål, der sigter mod episodisk genkaldelse, ikke altid udløser den tilsigtede genkaldelsesstrategi.

Omvendt er det også muligt, at spørgsmål, der sigter mod genkaldelse af informationer om det gængse eller "normale" og dermed mod aktivering af semantisk hukommelse, besvares ved aktivering af episodisk. Enkelte undersøgelser peger dog på, at sandsynligheden for akti-

vering af semantisk hukommelse ved besvarelse af episodiske spørgsmål er større end omvendt (se fx Jobe et al., 1993). Derimod synes det ofte forekommende, at respondenter aktiverer episodiske genkaldelsesstrategier ved besvarelse af holdningsspørgsmål, når holdninger er ikke-eksisterende.

Kognitive processer, der sigter imod genkaldelse af episodiske informationer, men som beror på aktivering af semantisk hukommelse, forekommer især, når der ønskes informationer om gængse hændelser, og/eller når den retrospektive tidshorisont er lang. Mangelfuld aktivering af episodisk hukommelse ved genkaldelse af sædvanlige hændelser i tid og rum synes navnlig at afhænge af hændelsers hyppighed (se fx Blair & Burton, 1987; Burton & Blair, 1991). Derfor anbefaler mange, at retrospektive spørgsmåls tidshorisonter reduceres (se fx Sudman et al., 1996).

Monosurveyisk metodeforskning sætter også fokus på processen *fra indkodning til genkaldelse af informationer*. Flere parametre påvirker genkaldelse af episodiske informationer, hvoraf én er manglende indkodning: Informationen er aldrig indkodet og kan derfor ikke genkaldes (se fx Fowler, 1995). Hvis informationen efterlyses, kan det føre til "genkaldelse" af aldrig indkodede informationer. Hvis informationen derimod er indkodet, er det ikke sikkert, at den findes. Der kan foreligge, hvad Jabine et al. (1984) kalder *simple failure to recall*. Informationen er gået til grunde i hukommelsen. Desuden kan informationen kontamineres med andre. Genkaldelse er kognitive processer, hvor der successivt "fyldes på", drages slutninger og tilføjes plausible detaljer (se fx Shum & Rips, i: Sirken et al., 1999).

Tidsafstande mellem indkodnings- og genkaldelsestidspunkt påvirker også genkaldelse, der ofte kræver tid, men ikke nødvendigvis er en simpel funktion af tiden fra indkodning til genkaldelse. Engang troede Ebbinghaus, at genkaldelse kunne beskrives ved en logaritmisk glemselskurve. Kurven forekom tilforladelig, da den svarer til, hvad mennesker erfarer. Imidlertid forsvarer kun få forskere Ebbinghaus' kurve og resultaterne af hans "selvforsøg" (se fx Eysenck, 1993; Eysenck & Keane, 1995). At forglemmelse ikke er simple virkninger af tidens gang skyldes bl.a., at visse informationer indkodes sikrere end andre. Informationer om emotionelt ladede begivenheder genkaldes ofte uden problemer. Derimod glemmes stereotype begivenheder ofte hurtigt (se fx Eysenck & Keane, 1995; Jabine et al., 1984).



I surveysammenhæng underbygges Ebbinghaus' glemselskurve empirisk i nogle tilfælde, men ikke i andre. Undertiden peges der på en negativ sammenhæng mellem tid og genkaldelse: Jo længere tidsmæssig afstand mellem indkodning og genkaldelse af hændelsesinformationer, desto mere mangelfuld genkaldelse (se fx Smith et al., 1991; Tanur, 1992). Andre undersøgelser udviser dog ikke sådanne tendenser (se fx Borrie & Roggenbuck, 1998). Der er derfor grund til at betvivle Ebbinghaus' glemselskurve, hvilket imidlertid ikke betyder, at respondenter ikke *ceteris paribus* har dårligere episodisk hukommelse, jo længere den tidsmæssige afstand er fra indkodning til genkaldelse. Pointen er, at alt andet sjældent er lige.

Yderligere en årsag til, at genkaldelse af episodiske informationer ikke er en simpel funktion af tiden fra indkodning til genkaldelse, er *teleskopering*. Sudman og Bradburn (1973) udviklede en teleskopisk forglemmelsesmodel. Modellen omfatter både simpel glemsel og teleskopering (tidskompression). Sammenpresningens virkning er, at hændelser placeres senere end det faktiske tidspunkt. Ved test af modellen peger Sudman og Bradburn (op.cit.; se også Sudman & Bradburn, 1974) på, at såvel glemsel som tidskompression spiller en rolle for episodisk genkaldelse. Modellen har dog siden været genstand for en del kritik (se fx Loftus et al., 1985) og forbigås nu i store dele af monosurveysk metode-litteratur.

I videre forstand kan teleskopering både ske fremad og bagud (se fx Sudman et al., 1996). Teleskopering underbygges empirisk, mens andre undersøgelser peger på moderate virkninger (se fx Braverman & Slater, 1996; Converse & Presser, 1986). I nogle undersøgelser vises, at hændelsers hyppighed øger sandsynligheden for mangelfulde tidsmæssige placeringer. Manglende karakteristiske træk ved begivenheder bevirker, at de placeres forkert på tidsaksen, men også karakteristiske træk kan føre til teleskopering (se fx Brown et al., 1985; Hippler et al., 1987). Flere forskere peger på, at korte retrospektive tidshorisonter øger sandsynligheden for teleskopering.

Nogle hændelsers *karakteristiske træk* er yderligere en årsag til, at episodisk genkaldelse ikke er en simpel funktion af tiden fra indkodning til genkaldelse. Uafhængig af tid genkaldes informationer med karakteristiske træk mere pålideligt (se fx Crano & Brewer, 2002). Emotionelle begivenheder huskes fx ofte glimrende (se fx Blair & Burton, 1987; Fowler, 1993; Groves, 1989; Jabine et al., 1984). Tilsvarende gælder andre be-

givenheder, der overskrider kognitive drejebøger: "... forgetting is related to elapsed time, salience and the number of events that compete with the particular event(s) that the respondent is asked to remember" (Foddy, 1993:93). Vendepunkter og emotionelt ladede begivenheder er kognitive drejebøgers modsætning. Derimod glemmes stereotype begivenheder ofte hurtigere end usædvanlige, hvilket allerede Sudman & Bradburn (1973) gjorde opmærksom på, om end de ikke inddrog parameteren i deres model.

Nogle respondenters genkaldelse hæmmes af kognitive belastninger. Visse spørgsmål er vanskelige og kræver genkaldelse af komplekse informationer. Hertil kommer, at nogle er mere øvede end andre. For mindre øvede kan virkningen blive *kognitiv magelighed*, dvs. "genkaldelse" af tilfældige informationer (se fx Braverman, i: Braverman & Slater, 1996; Krosnick, 1991; Krosnick & Alwin, 1987; Loftus et al., i: Tanur, 1992; Schuman & Presser, 1981). Undersøgelser peger på, at mindre øvede ofte afgiver "omkostningsfrie" svar. Det gælder især holdnings-spørgsmål, hvor respondenter undertiden har "holdninger" til fiktive fænomener eller *nonattitudes* i form af tilsyneladende "holdninger" (se fx Schuman & Presser, 1996; Clark & Schober, i: Tanur, 1992). Pseudogenkaldelse kaldes undertiden *satisficing*, der afhænger af tre parametre: "... the first is the inherent *difficulty of the task* that the respondents confronts; the second is the respondent's *ability* to perform the required task; and the third is the respondent's *motivation* to perform the task" (Krosnick, 1991).

Eksempler på *kognitive aflastninger* er allerede påpeget, men monosurvey-sk metodelitteratur indeholder yderligere eksempler. Én overordnet anbefaling er, at "gode" spørgeskemaer forudsætter, at forskere har betydelig indsigt i, hvordan den menneskelige hukommelse fungerer (se fx Foddy, 1993; Krosnick, i: Braverman & Slater, 1996; Sudman et al., 1996; Tourangeau et al., 2000). Navnlig CASM-bevægelsen understreger dette som en forudsætning for at kunne reducere spørgsmåls kognitive sværhedsgrad.

Fremstillingen vedrørende episodisk og generel hukommelse viser, at der ikke altid er kongruens mellem spørgsmåls intenderede og respondenters faktiske genkaldelse. Én anbefaling er derfor, at spørgsmål formuleres på måder, der aktiverer relevante genkaldelsesstrategier. En anden anbefaling er, at spørgsmåls retrospektive tidshorisonter reduceres mest muligt (se fx Nolinke, 1998). Yderligere en anbefaling angår hæn-

delsers karakteristiske træk. Da gentagne analoge hændelser øger sandsynligheden for mangelfuld genkaldelse, bør dette indtænkes. Hvad holdningsspørgsmål angår, anbefaler mange – fx med henvisning til *non-attitudes* – anvendelse af ved ikke-svarkategorier, der ikke løser *satisficing*-problemet, men som i nogen grad kan forebygge måleproblemer.

De hidtidige eksempler viser, at vilkår for genkaldelse *per se* spiller en rolle for, om genkaldelse fører til måleproblemer. Informationer om hændelser kontamineres undertiden med aktuelle refleksioner. Ikke kun det indkodede og måden, hvorpå det er indkodet, spiller en rolle. Følgelig kan episodisk genkaldelse fremmes af vink (*cues*), især hvis de er beslægtet med omstændigheder under indkodningen (se fx Shum & Rips, i: Sirken et al., 1999). En type af vink er *landmarks*, dvs. tidlige orienteringspunkter (se fx Tanur, 1992; Tourangeau et al., 2000). Mens generelle orienteringspunkter er tidspunkter for begivenheder, som alle formodes at have kendskab til, er specifikke orienteringspunkter væsentlige hændelser, som den enkelte respondent har oplevet, fx ægteskab, barns fødsel eller skilsmisse.

Andre kognitive vink er *aided recall*, *bounded recall* og *record checks* (se fx Dijkstra & Van der Zouwen, 1982; Rossi et al., 1983). Genkaldelse kan fremmes ved, at interviewere læser svarkategorier op, dvs. *aided recall*. Hvis respondenter deltager i en longitudinel survey, kan genkaldelse fremmes ved at gentage oplysninger, som respondenter tidligere har afgivet, dvs. *bounded recall*. Endelig kan også forskellige former for dokumentation – *record checks* – virke fremmende for genkaldelse af informationer, fx selvangivelser, dagbøger, lejekontrakter o.l.

Yderligere aflastninger kan stimulere genkaldelse. I nogle undersøgelser peges på en sammenhæng mellem genkaldelse og den svartid, som respondenter tildeles. Jo mere tid, desto gunstigere er udsigten til pålidelig genkaldelse (se fx Schwarz et al., 1991). Måske derfor anbefaler flere forskere lange spørgsmål (se fx Fowler, i: Bickman & Rog, 1998). Genkaldelse kan ligeledes stimuleres af flere spørgsmål om samme anliggende.

### **Multisurveysk metodeforskning**

Efter tematisering og kondensering af monosurveysk metodeforskning med fokus på respondenters genkaldelse af informationer bringes om muligt empiriske resultater fra tilsvarende multisurveysk forskning på bane. Med forbehold for eventuelle litteraturheuristiske mangler viser det

sig imidlertid, at multisurveysk forskning er tavs om de fleste af de i *oversigt 4.3.* nævnte monosurveyske aspekter. Kun om strategier for genkaldelse af informationer samt kognitiv magelighed findes enkelte resultater. Hvad kognitiv magelighed angår, drejer det sig især om føjelighed (se *oversigt 4.4.*).

#### **Oversigt 4.4.**

Aspekter af genkaldelse ifølge multisurveysk metodeforskning.

---

- Genkaldelsesstrategier
  - Føjelighed (*acquiescence*)
  - På midten- og pol-svar (se afsnit 4.2.)
- 

Hvad *genkaldelsesstrategier* angår, er det som nævnt afgørende, at informationer genkaldes som forskningsmæssigt tilsigtet, fx at episodiske informationer genkaldes episodisk og semantiske ved aktivering af generel hukommelse. Om denne problematik findes enkelte multisurveyske undersøgelser, der samtidig sætter fokus på andre svarprocessuelle aspekter, fx forståelse af spørgsmål (se fx Ji et al., 2000; Johnson et al., 1997; Raitasalo et al., 2005; Warnecke et al., 1997).

En tidligere nævnt amerikansk undersøgelse handler om genkaldelsesstrategier blandt amerikanere med afrikansk, mexicansk, puertoricansk og europæisk baggrund, blandt hvilke forskerne forventer at finde kulturelt betingede forskelle: "Because of the often overwhelming amount of information that individuals must process as they navigate their environment, cultural conditioning and values are likely to influence the subset of information that is identified as important and encoded" (Johnson et al., 1997:90). Baseret på *respondent-probing* findes imidlertid ingen kulturspecifikke forskelle med hensyn til semantisk genkaldelse af informationer om hyppige begivenheder og episodisk genkaldelse af sjældnere begivenheder, fx lægekonsultationer. Derimod finder Johnson et al. forskelle med hensyn til genkaldelse af episodiske informationer: "African-Americans were more likely to prefer free order of recall strategies when conducting an episodic memory search. Most respondents from other cultural groups preferred more linear (either forward or backward) search strategies" (op.cit.:107). Ifølge Johnson et al. er undersøgelsen et beskedent bidrag til forskning om kulturelle genkaldelsesfor-

skelle. Forskerne beklager den begrænsede forskning herom og efterlyser yderligere metodeforskning.

En anden undersøgelse sætter komparativt fokus på genkaldelsesstrategier blandt kinesere og amerikanere (Ji et al., 2000). Forskerne tager afsæt i begreberne kollektivism versus individualisme og fastslår, at kollektivistiske kulturer er orienteret imod social afhængighed og integration, mens individualistiske kulturer er ensbetydende med uafhængighed og autonomi. Disse kulturelle forskelle i selvets konstruktion reflekteres i menneskers selvbiografiske hukommelse. På den baggrund sætter Ji et al. (op.cit.) fokus på forskelle mellem kineseres og amerikaneres genkaldelse af informationer, fx om så forskellige fænomener som antal biblioteksbesøg og forkølelser etc. inden for specificerede perioder. Såfremt kinesere i højere grad end amerikanere er opmærksomme på såvel egen som andres adfærd, bør førstnævnte i mindre grad end sidstnævnte genkalde informationer som estimeret genereret ved hjælp af svarkategorier, hvilket er, hvad forskerne finder: "Chinese respondents ... were less influenced by the response alternatives than were Americans for observable behaviors, both in their behavioral reports and comparative judgments" (Ji et al., 2000:585).

Med disse analyseresultater antydes, at mennesker tilhørende kollektivistiske kulturer indkoder mere detaljerede informationer angående almindelig hverdagsadfærd end mennesker fra individualistiske: "Collectivist cultures put a premium on fitting in, which requires considerable monitoring of both one's own behaviour and that of others ..." (Ji et al., 2000:591). Det er en problematik, som også andre forskere er optaget af. I et litteraturstudie reflekterer Uskul og Oyserman (2005) med afsæt i kollektivism og individualisme over mulighederne for at overføre CASM-bevægelsens monosurveyske landvindinger til multisurveys. Forskerne gør gældende, at kulturelt tilhørsforhold ikke kun påvirker forståelse af spørgsmål, men også genkaldelse af informationer: "Collectivism increases sensitivity to situation, influences what is stored and accessible in memory, subjective theories and what is socially desirable and requires extremity or modesty. Collectivism also influences what is salient enough to be memorable and not require estimation" (Uskul & Oyserman, 2005:28).

Blandt fåtallige eksempler skal resultater fra en tidligere nævnt finsk undersøgelse baseret på kognitive interview fremhæves (Raitasalo et al., 2005). Når det handler om finners, tyskeres, hollænderes og italiene-

res genkaldelse af informationer om alkoholvaner, fx deres hyppighed, mængde og brugssteder, er der kun få genkaldelsesstrategiske variationer: "... the overall structure is remarkably similar in the four countries" (op.cit.:367). Når hertil lægges resultater fra Warnecke et al.'s (1997) førnævnte Chicago-undersøgelse, der på samme måde udviser ubetydelige kulturelt betingede genkaldelsesstrategiske variationer, forbliver det uafklaret, hvornår og under hvilke betingelser genkaldelsesstrategier måtte være kulturelt sensitive. Som bl.a. påpeget af Johnson et al. (1997) er det en så godt som uudforsket problematik med et betydeligt forskningsbehov, der fx vil kunne tilgodeses ved at overføre CASM-erfaringer til multisurveysk metodeforskning.

Multisurveysk metodeforskning er lidt mere udbredt, hvad undersøgelse af føjelighed (*acquiescence*) angår.<sup>47</sup> Føjelighed, der ligesom vekslende strategier til genkaldelse af informationer kan skabe ækvivalensproblemer, er som nævnt beslægtet med kognitiv magelighed, bl.a. fordi begge er kendetegnet ved bagvedliggende *satisficing*-strategier: "... a great deal of evidence (vedr. føjelighed, h.o.) is consistent with satisficing and cannot be accounted for by ... other explanations" (Krosnick, 1999).<sup>48</sup> Adskillige undersøgelser sætter fokus på respondenters nationalt og/eller kulturelt bestemte føjelighed, der undertiden kaldes *yea-saying bias*, dvs. en tendens til at erklære sig enig i ethvert udsagn, uanset hvad der måtte blive spurgt om (se fx Baumgartner & Steenkamp, 2001; Cheung & Rensvold, 2000; Harzing, 2006; Javeline, 1999; Johnson et al., 1997; Johnson et al., 2005; Smith, 2004; Smith, 2004a; Van Herk, 2000; Van Herk et al., 2004).<sup>49</sup>

Metodeforskningens gennemgående tendens er, at føjelighed eller ja-sigen, der navnlig forekommer ved besvarelse af holdningsspørgsmål, genererer måleproblemer, som ikke blot bør tages ad notam, men om muligt forebygges: "Over the past 2 decades, the majority viewpoint

---

47. I dele af metodelitteraturen nævnes føjelighed og afgivelse af pol-svar (se afsnit 4.2.) som eksempler på, hvad der kaldes *response styles* (se fx Harzing, 2006).

48. "For example, acquiescence is more common among people with more limited cognitive skills (...) and with less cognitive energy (...), and among those who do not like to think (...). Acquiescence is more common when a question is difficult to answer (...), when respondents have been encouraged to guess (...), after they have become fatigued (...), and during telephone interviews than during face-to-face interviews (...), presumably because people feel more accountable under the latter conditions. People who acquiesce are likely to manifest other forms of satisficing (...), such as nondifferentiation (...) and selecting a no-opinion option" (Krosnick, 1999).

49. Se fx en omfattende forskningskondensering i: Harzing, 2006.

among cross-cultural researchers has been that acquiescent response bias is a source of error in the making of cross-cultural comparisons and that ways must thus be found to discount it" (Smith, 2004a:51). Forebyggelse kan fx ske ved reversering af svarkategorier, som det fx er velkendt ved konstruktion af Likert-skalaer (se fx Smith, i: Presser et al., 2004a).

At respondenters føjelighed ikke er et ækvivalensproblem, der bør abstraheres fra, fremgår af følgende eksempler: En undersøgelse peger på væsentlige forskelle med hensyn til føjelighed blandt respondenter i Kasakhstan (Javeline, 1999). Tendens til ja-sigen, der så at sige skjuler "sande" holdninger, forekommer i hele stikprøven (N = 1.986), men viser sig at være markant større blandt etniske kasakhstanere end blandt etniske russere: "This analysis demonstrates that there is a widespread tendency for individuals to agree with statements that run counter to their true preferences" (op.cit.:23f). Det hermed skabte ækvivalensproblem foreslås løst ved formulering af balancerede spørgsmål kombineret med mere substantielle svarkategorier end *enig/ ikke enig*.

Baseret på data om madlavning, opvask og personlig pleje i Grækenland, Italien, Spanien, Frankrig, Tyskland og UK viser en tidligere nævnt undersøgelse, at føjelighedstendenser er mere udbredt i lande omkring Middelhavet – navnlig blandt grækere – end i det nordvestlige Europa (Van Herk et al., 2004; se også Van Herk, 2000). I yderligere en undersøgelse omfattende 26 lande peges ligeledes på markante forskelle bl.a. med hensyn til ja-sigen: "Our study has shown that there are substantial differences in response styles across countries that ... confirmed patterns found in earlier studies" (Harzing, 2006:257). Sidstnævnte undersøgelse omhandler også andre "stilistiske" forskelle, fx tendenser til pol-svar.

I en undersøgelse, der sætter fokus på respondenters trusselsfølelser i forhold til indvandrere i ni vesteuropæiske lande, konkluderes ligeledes, at føjelighed eksisterer: "The results clearly show that in each of the countries the scale that measures the respondent's feelings of threat towards immigrants is susceptible to an agreeing-response bias" (Billiet et al., 2002:69).<sup>50</sup> Data menes påvirket af en uønsket kulturelt

---

50. Artiklens formål er at vurdere omfanget af konceptuel ækvivalens med hensyn til operationalisering af trusselsfølelser og baseres på *structural equation modelling*: "The introduction of a style factor allows us to control for a possible source of measurement non-equivalence, namely method bias" (Billiet et al., 2002:53).

betinget metodeffekt, men i dette tilfælde dog i nogenlunde samme omfang i alle ni lande. Analyserne viser også, at en model, der inkluderer en "svarstil"-faktor altid passer bedre på data end modeller uden en sådan (op.cit.).<sup>51</sup>

Når det handler om at forklare fænomenet føjelighed, peger en undersøgelse på, at det er mere udbredt i kollektivistiske end individualistiske samfund (Smith, 2004a). Uafhængighed og autonomi kan med andre ord forebygge føjelighed. På grundlag af stikprøver fra 19 lande konstaterer også Johnson et al. (2005), at individualisme hæmmer tendenser til føjelighed: "Individualism, uncertainty avoidance, power distance, and masculinity were each found to be negatively associated with acquiescent response behaviour" (p. 264).<sup>52</sup> En anden forklaring findes hos Cheung og Rensvold (2000), der peger på respondenteres opfattelser af, hvad der anses for at være "socialt ønsket" som afgørende (se afsnit 4.5.): "Cross-cultural differences in ARS (dvs. acquiescence response styles, h.o.) can be explained in terms of social desirability, a belief that a higher score is a better score, or by preoccupation with individual defects and deficiencies (...). On the individual level, some respondents display extreme ARS by agreeing (or disagreeing) with almost any statement" (op.cit.:190). Som andre forskere pointerer også Cheung og Rensvold (op.cit.), at føjelighed kan generere spuriøse resultater, der er langt fra "sande" holdninger og holdningsforskelle.

#### 4.4. KONTEKSTEFFEKTER

Der er hidtil abstraheret fra, at spørgsmålsforståelse og genkaldelse kan afhænge af den sammenhæng, hvori spørgsmål findes. Ikke kun forståelse af ord, spørgsmål og svarkategorier fører til måleproblemer. Det gæl-

---

51. I en marketingsundersøgelse omfattende 11 EU-lande findes ligeledes tendenser til føjelighed og andre "stilistiske" svarstrategier, fx pol-svar: "Response styles are a source of contamination in questionnaire ratings, and therefore they threaten the validity of conclusions drawn from marketing research data" (Baumgartner & Steenkamp, 2001).

52. "Data are presented from approximately 18,000 survey questionnaires completed by employees in 19 nations on five continents (Australia, Belgium, Brazil, Czech Republic, Germany, Hungary, India, Japan, Malaysia, Portugal, Turkey, the United Kingdom, Mexico, the Philippines, Poland, Singapore, Hong Kong, France, and Italy). Hierarchical linear modeling was employed to examine the associations between person-level response styles and country-level cultural orientations" (Johnson et al., 2005; uddrag af abstract).



der også konteksten, der som kognitiv overførsel to eller flere spørgsmål imellem kan generere konteksteffekter. Konteksten kan ikke blot påvirke forståelse, men alle faser i svarprocessen: "The context in which a question is asked can influence respondents' cognitive processes at each stage of the question answering process. Moreover, the various processes are not mutually exclusive; a given context effect may reflect the simultaneous operation of several processes" (Sudman et al., 1996:98f). Det er, hvad dette afsnit handler om. Som tidligere sættes der både fokus på mono- og multisurveysk metodeforskning, hvor sidstnævnte viser sig at være sporadisk.

### **Monosurveysk metodeforskning**

Konteksteffekter har været kendt længe, men der er ikke enighed om, hvorfor og hvor ofte den slags effekter fører til måleproblemer (se fx Smith, i: Biemer et al., 1991). Konteksteffekter indebærer, at spørgsmålsforståelse og andre kognitive processer afhænger af den sammenhæng, hvori spørgsmål stilles og besvares: "... question comprehension is not merely a function of the wording of the question itself. Respondents use information provided by the context of the question to determine its intended meaning" (Sudman et al., 1996:69). Spørgsmål findes i en klassisk del/helhed-kontekst. Spørgsmål forstås og svar afgives inden for kontekstens ofte tavse baggrundsforståelse (se fx Fienberg & Tanur, 1989; Schwarz & Sudman, 1992; Tanur, 1992).

Amerikansk og anden monosurveysk metodeforskning sætter fokus på spørgsmålskontekster som årsag til måleproblemer. Fokuseringen er ikke overraskende, eftersom konteksteffekter har vist sig at bevirke, at respondenter besvarer "samme" spørgsmål forskelligt i forskellige kontekster. Konteksteffekter kan derfor bevirke, at et spørgsmål, der fører til valide målinger i én sammenhæng, ikke gør det i en anden (se fx Schwarz & Sudman, 1992). Det "gode" spørgsmål findes altså ikke uafhængigt af den sammenhæng, det indgår i. Metodeforskningen sætter fokus på forskellige kontekstuelle aspekter (se *oversigt 4.5.*).

## Oversigt 4.5.

Kontekstuelle aspekter ifølge monosurveysk metodeforskning.

Generelle aspekter	Typologier
Udbredelse	Definitionseffekt
Spørgsmålstyper	Redundanseffekt
Sammenhængende spørgeforløb	Simpel kontrasteffekt
	Subtraktionseffekt
	Fokuseringseffekt
	Stimuleringsseffekt

Selv om monosurveysk metodeforskning er mangelfuld, bl.a. fordi den ikke udpeger sikre metoder til reduktion af konteksteffekter, er effekter af forskellig type påvist. Allerede i begyndelsen af 1990'erne lokaliserede Schwarz & Sudman (1992) over 25 undersøgelser, hvori den slags effekter påvises (se fx Drew, 1991; Harrison & McLaughlin, 1993; Schuman & Presser, 1981; Strack et al., i: Schwarz & Sudman, 1992). Siden er adskillige undersøgelser føjet til listen (se fx McLaughlin, 1999; Schuman & Presser, 1996; Tourangeau, i: Sirken et al., 1999). Der findes dog også undersøgelser, hvor konteksteffekter ikke forekommer eller er moderate (se fx Smith, i: Biemer et al., 1991). Blandt andet derfor er det vanskeligt at forudsige konteksteffekters omfang. Undertiden bidrager konteksten ligefrem til reduktion af måleproblemer, dvs. genererer *positive* konteksteffekter, fx hvis spørgeforløb støtter respondenter i forståelse eller informationsgenkaldelse.

Alt i alt underbygger monosurveysk metodeforskning, at konteksteffekter er årsag til måleproblemer. Konteksteffekter synes navnlig at forekomme i tilknytning til besvarelse af generelle holdningsspørgsmål (se fx Bradburn, i: Schwarz & Sudman, 1992). Tourangeau (i: Sirken et al., 1999) forklarer tendensen med, at holdninger i højere grad end faktualiteter er skrøbelige *ad hoc*-konstruktioner og særligt udsatte for konteksteffekter. Uddybende om konteksteffekter i tilknytning til besvarelse af holdningsspørgsmål påpeger Tourangeau et al. (2000): "The question's context can affect what respondents consider as they answer the question; in addition, context can, by suggesting a standard of comparison, alter how respondents evaluate what comes to mind" (Tourangeau et al., 2000:229).

Konteksteffekter er dog også påvist ved besvarelse af faktuelle spørgsmål (se fx Baldursson & Pedersen, 1992). I en anden undersøgelse forsøgte Presser (1990a) at reducere amerikaneres overrapportering af valgdeltagelse ved at ændre konteksten. Der kan endvidere forekomme kontekstuel betingede sammenhænge mellem besvarelse af holdnings-spørgsmål og faktuelle spørgsmål (se fx Converse & Presser, 1986).

Konteksteffekter synes især at forekomme ved besvarelse af tematisk sammenhængende spørgeforløb. Flere forskere peger på, at sandsynligheden for konteksteffekter er størst, når holdningsspørgsmål er tematisk beslægtede (se fx Houtkoop-Steenstra, 2000; Tourangeau et al., 2000). Hvis fx et specifikt holdningsspørgsmål placeres før et generelt, kan svaret på det specifikke smitte af på det generelle (*carryover effect*). Overførsel finder sted, når spørgsmål danner en forståelsesramme, som påvirker efterfølgende svar: "Carryover effects occurs when the interpretation, retrieval, judgment, or response associated with prior items provides a respondent with an easily accessible cognitive structure" (Harrison & McLaughlin, 1993).

En konteksteffekt er *definitionseffekt* (se fx Smith, i: Schwarz & Sudman, 1992). Effekter af den type skyldes kohærensøgende forståelse af spørgsmål i spørgeforløb. Hvis fx et spørgsmål om kommunistiske journalisters ret til (før USSR's sammenbrud) at arbejde i USA stilles lige efter et spørgsmål om amerikanske journalisters ret til at arbejde i kommunistiske lande, medfører definitionseffekt, at emnet for det sidste spørgsmål er holdninger til ytringsfrihed, ikke til kommunister. Også *redundanseffekter* kan påvirke forståelsesprocessen, når flere spørgsmål angår aspekter af samme emne. Hvis respondenter ikke forventer redundante spørgsmål, ekskluderes tidligere aspekter fra forståelsen (se fx Bradburn, 2000).

*Simpel kontrasteffekt* findes, når respondenter tilsidesætter personlige præferencer og forholder sig til et emne i modsætning til tidligere (se fx Tourangeau et al., 2000). Hvis et generelt spørgsmål om ret til fri abort stilles efter et spørgsmål om ret til abort efter voldtægt, kan effekten bevirke, at nogle respondenter udtrykker større forbehold over for det generelle spørgsmåls emne, end hvis det var stillet alene.

Yderligere en effekt er *subtraktionseffekt*. Når respondenter tilstræber at afgive nye informationer for hvert nyt spørgsmål og ikke gentager tidligere afgivne, men relevante informationer, foreligger der subtraktionseffekt (se fx Se fx Smith, i: Schwarz & Sudman, 1992; Tou-

rangeau et al., 1991). Når et spørgsmål om ægteskabelig lykke efterfølges af et spørgsmål om lykke i al almindelighed, trækker nogle respondenter den ægteskabelige lykke fra ved besvarelse af det generelle spørgsmål og udtrykker dermed en mindre grad af generel lykke, end hvis spørgsmålet var blevet besvaret først.

Også *fokuseringseffekter* er påvist (se fx Smith, i: Schwarz & Sudman, 1992). Når respondenter ved besvarelse af førnævnte generelle spørgsmål om lykke tager afsæt i dét aspekt af den generelle lykke, som var i fokus i det forudgående spørgsmål, er der fokuseringseffekt. Nogle respondenter påvirkes også af *stimuleringsseffekter*, der kan forekomme, når informationer, der netop er genkaldt, fremmer genkaldelse af yderligere informationer om samme emne. Hvis effekten fx resulterer i overrapportering fx på grund af kognitiv magelighed, opstår der måleproblemer. Hvis forudgående spørgsmål derimod støtter hukommelsen, reduceres måleproblemer.

Uanset konteksteffekters art er det et problem, at der ikke er udviklet effektive procedurer til forudsigelse og forebyggelse af kontekstuelle måleproblemer: "There seems to be no general theory that predicts when such effects are to be expected and when they should not be expected" (Groves, i: Biemer et al., 1991). Nogle forskere peger dog på en række anbefalinger, som kan reducere konteksteffekter, fx følgende: "In general, the question or set of questions most vital to the research should be presented first, to avoid unwanted influence from preceding questions, due to their effects on information accessibility and current question interpretation" (Tourangeau, i: Sirken et al., 1999). Der peges også på moderat anvendelse af tematiske spørgeforløb, ligesom det understreges, at test af spørgeskemaer kan bidrage til reduktion af negative konteksteffekter.

### **Multisurveysk metodeforskning**

Den forholdsvis righoldige metodeforskning om konteksteffekter i monosurveys – herunder effekternes vekslende udbredelse og former – viser sig ikke i nævneværdigt omfang at have afsmittende virkning på tilsvarende multisurveysk forskning, hvoraf så godt som ingen synes eksisterende. Når konteksteffekter skaber måleproblemer i monosurveys, bør antagelsen fx ifølge Smith (i: Harkness et al., 2003) være, at effekterne næppe er af mindre omfang, når grænser eller kulturer overskrides. Ikke desto mindre er det med litteraturheuristiske reservationer kun lyk-

kedes at lokalisere et par empiriske undersøgelser, der sætter fokus på konteksteffekter (se Braun, i: Harkness et al., 2003; Schwarz, i: Harkness et al., 2003).

I en undersøgelse, der også handler om, hvorvidt respondenter påvirkes af svarkategorier, får tyske og kinesiske studerende stillet kontekstuel reverserede spørgsmål om studiemæssig tilfredshed og tilfredshed i almindelighed med det udfald, at afgivne svar afhænger af spørgsmålenes rækkefølge (Schwarz, i: Harkness et al., 2003). Når spørgsmålet om studietilfredshed stilles først, viser der sig væsentlige forskelle mellem tyskere og kinesere, mens det modsatte er tilfældet, når spørgsmålenes rækkefølge reverseres: "... the German respondents drew on the information brought to mind by the academic question when they subsequently evaluated their general life satisfaction" (op.cit.:98).

I en anden undersøgelse, der sætter fokus på kulturelt betingede konteksteffekter, peger Braun (i: Harkness et al., 2003) på det fra monosurveyforskning velkendte fænomen, at respondenter inddrager alt, hvad de opfatter som relevant, for at kunne besvare spørgsmål, herunder også konteksten, der både omfatter spørgsmålsrækkefølge og kontekst i bredere betydning. Denne andet-inddragelse gælder navnlig, når spørgsmål har stor betydningvidde og respondenter derfor får brug for "hjælp" andetsteds. Alene derfor kan konteksteffekter, påpeger Braun (op.cit.), være meget forskellige på tværs af landegrænser, og citerer derpå Clark og Schober (1992): "If the surveyer thinks this word has an obvious meaning, then it must be the meaning that is obvious to me at the moment" (p. 28).

Trods marginal empirisk baseret fokusering på konteksteffekter er det et problem, som adskillige multisurveyforskere er opmærksomme på og gerne ser forebygget (se fx Braun, i: Harkness et al., 2003; Saris, 1998; Schwarz, i: Harkness et al., 2003; Smith, i: Harkness et al., 2003; Smith, i: Presser et al., 2004a; Usunier, 1999). Da konteksteffekter er velkendte og synes at forekomme i de fleste monosurveys, er det fx efter Smiths opfattelse beklageligt, at metodeforskningen stort set er tavs, når det handler om "grænseoverskridende" undersøgelser af konteksteffekter: "Research indicates that survey context effects operate in most countries, but experiments and detailed studies have been conducted in only a small number of countries and coordinated, cross-national, experimental studies are lacking" (Smith, i: Harkness et al., 2003:84).

#### 4.5. SVARREDIGERING

Selv hvis det utænkelige skulle ske, at hverken forståelse af spørgsmål, genkaldelse eller konteksten, hvori spørgsmål findes, genererer måleproblemer, er respondenteres *svarredigering* yderligere en årsag til måleproblemer. Respondenter kan have genkaldt valide informationer, som de ikke ønsker formidlet, fx fordi de opfatter informationerne som ”private”, sensitive eller potentielt stigmatiserende. Som tidligere sættes der først fokus på monosurveysk metodeforskning, hvorpå opmærksomheden rettes imod multisurveysk, der omfatter adskillige undersøgelser.

##### **Monosurveysk metodeforskning**

Monosurveyske metodeforskere peger på, at respondenter ofte – under iagttagelse af *social desirability* og *self presentation* – redigerer genkaldte informationer (se fx Fowler, i: Bickman & Rog, 1998; Jobe et al., i: Lyberg et al., 1997; Schwarz, i: Lyberg et al., 1997; Sudman et al., 1996; Tourangeau et al., 2000). Det nævnes også, at svarredigering ikke kun angår selvpresentation af ”socialt ønskede” adfærdsformer og holdninger, men er redigering i videre forstand, dvs. formidling af, hvad respondenter opfatter som ”passende” svar. Nogle respondenter vurderer fx genkaldte informationers tilstrækkelighed omhyggeligt, før svar afgives, mens andre er mindre pertentlige. Hvad besvarelse af holdningsspørgsmål angår, kan redigering spille en afgørende rolle for måleproblemer, påpeger Tourangeau et al. (op.cit.), da respondenter undertiden tilbageholder eller modificerer konstruerede holdninger, når disse opleves som utilstrækkelige.

Blandt nogle respondenter opfattes visse spørgsmål som sensitive. I monosurveysk metodeforskning er der dog ikke konsensus om, hvad sensitive spørgsmål er. Når forskere fx ønsker informationer om alkoholisme, brug af hash eller hårde stoffer, abort, seksuel eller kriminel adfærd, nærmer forskere sig temaer, der er følsomme for mange, fordi spørgsmålene angår socialt uønskede tilstande og adfærdsformer. Sensitive spørgsmål, påpeger Sudman og Bradburn, er “... largely personal and directly threatening to the respondent's self-esteem or are anxiety-provoking” (Sudman & Bradburn, 1974). Når spørgsmål omhandler den slags anliggender, tilbydes undertiden ”socialt ønskede” svar (se fx Groves, 1989). Selv om spørgsmål forstås og relevante informationer genkaldes, afgives svar, der afspejler iagttagelse af fremherskende sociale nor-

mer, og som fx også afhænger af interaktioner mellem respondenter og interviewere (se fx Sudman et al., 1996).

Med hensyn til *social desirability*-baseret redigering er den mono-surveyske tendens, at respondenter bestræber sig på at fremstille sig selv som sympatiske: ”Studies of response accuracy suggest that there is a tendency among respondents to distort answers in ways that will make them look better or will avoid making them look bad” (Fowler, i: Bickman & Rog, 1998:354). Ifølge Fowler er problemet ikke, at visse spørgsmål er ”sensitive”, men at svar kan være det (op.cit.). Spørgsmål om brug af rusmidler er fx næppe sensitive for ikke-brugere, men er det til gengæld blandt mange storbrugere. Der er to mulige virkninger af den slags redigeringer, nemlig at hyppigheden af ”socialt uønsket” adfærd generelt undervurderes, mens omvendt hyppigheden af ”socialt ønsket” overvurderes (se fx Tourangeau et al., 2000). Over- og undervurderinger afdækkes almindeligvis ved at sammenholde surveydata med valide registerdata.

Respondenters over- og undervurdering af ”socialt ønskede” tilstande og adfærdsformer kan være af marginal betydning, men er ofte markant: ”Survey respondent’s decision to edit their answers can have a big impact on the estimates from surveys” (Tourangeau et al., 2000:287). Hvad undervurdering angår, fandt Bradburn og Sudman (1979) fx markante frekvensmangler om motorkørsel i beruset tilstand. Senere viste Leinhardt (1981), at surveymetoden er uegnet til at belyse forekomst af kønssygdomme og andre seksuelt overførte sygdomme, eftersom frekvenser underrapporteres markant. I yderligere en undersøgelse om amerikaneres helbredsadfærd findes en markant undervurdering af lægebesøg, mens antallet af specifikke behandlinger – det gælder fx vaccinationer – overvurderes (Loftus et al., i: Tanur, 1992; tilsvarende, se fx Tourangeau et al., 2000).

Sættes der vice versa fokus på ”socialt ønskede” tilstande og adfærdsformer, peger undersøgelser på større eller mindre overvurderinger, dog normalt ikke i samme omfang som det ”socialt uønskede” undervurderes. Eksempelvis oplyser amerikanere at have deltaget i valg handlinger i et omfang, der signifikant overskrider faktisk stemmeafgivelse (se fx Kalton & Schuman, 1982; Tanur, 1992; Presser, 1990a; Tourangeau et al., 2000). Den tidligere nævnte overvurdering af gennemførte vaccinationer og andre lægebehandlinger er yderligere et eksempel (Loftus et al., i: Tanur, 1992). I Danmark synes respondenter fx at overvurdere hyppig-

heder, når de spørges om, hvor ofte de læser politiske nyheder i dagblade eller ser tv-avis.<sup>53</sup>

Monosurveysk metodeforskning peger på forskellige procedurer til reduktion af over- og underrapportering. Én anbefaling er sikring af anonymitet og fortrolighed, idet undersøgelser peger på, at målinger af socialt uønskede tilstande og adfærdsformer bliver mere valide, når respondenter ikke drager deres deltagelsesanonymitet i tvivl (se fx Fowler, 1995; Fowler, i: Bickman & Rog, 1998; Schwarz et al., 1991; Tourangeau et al., 2000). Yderligere en anbefaling er at understrege vigtigheden af nøjagtige svar: "... make clear to respondents that being accurate is more important than self-image or rapport with the interviewer" (Fowler, 1995:45).

Enkelte undersøgelser tyder på, at åbne spørgsmål er at foretrække ved besvarelse af sensitive spørgsmål (se fx Groves, 1989). Imidlertid synes datatilvirkningsformen mere afgørende, hvor postspørgeskemaer er at foretrække. Sudman et al. (1996) finder fx, at flere er "lykkelige", når de besøgsinterviewes, end når datatilvirkning beror på selvudfyldelse af postspørgeskemaer. Det er ikke overraskende, at redigerende overvejelser om det "socialt uønskede" spiller en særlig rolle, når interviewere varetager tilvirkningen af data. På den anden side kan besøgsinterview være at foretrække frem for telefoninterview. Fowler (1993) peger fx på, at besøgsinterview virker mere over- og underrapporteringsreducerende end telefoniske. Ingen interviewform løser dog problemet fuldt ud. Sudman og Bradburn (1996) påpeger fx, at ingen måde, hvorpå data tilvirkes, er optimal for opnåelse af valide svar på sensitive spørgsmål.

Ud over de nævnte anbefalinger findes en særlig procedure, hvorigennem "socialt uønsket" adfærd kan afdækkes til forebyggelse af måleproblemer. Proceduren – *randomized response technique* – indebærer, at interviewere ikke informeres om, hvad respondenter svarer på: "The technique introduces explicit randomization at the person level so that some respondents are asked the sensitive question and some other are asked another question. The choice of which question is asked is left to a box of colored balls or some other method" (Groves, 1989). Groves og andre peger på, at proceduren fører til færre måleproblemer (se fx Fowler, i: Bickman & Rog, 1998). Som én svaghed påpeges dog, at *randomized*

---

53. Personlig oplysning v/prof. Tom Bryder, Växjö Universitet.



*response*-proceduren fordrer forholdsvis store stikprøver. Hertil kommer, at respondenter selvsagt skal være overbeviste om tilsikret anonymitet, for at proceduren virker efter hensigten.

### **Multisurveysk metodeforskning**

Efter summarisk kondensering af monosurveysk forskning med fokus på respondenteres redigering af genkaldte informationer bringes resultater fra tilsvarende multisurveysk forskning på bane. Inden for denne problematik, der i de fleste tilfælde anskues som én af flere former for *response style*, er forskningen mindre tavs end vedrørende konteksteffekter (se fx Baumgartner & Steenkamp, 2001; Johnson et al., 1997; Johnson et al., 2001; Johnson & Van de Vijver, i: Harkness et al., 2003; Lalwani et al., 2006; Middleton & Jones, 2000; Ross & Mirowsky, 1984; Smith, 2004; Smith, i: Presser et al., 2004a; Stocké, 2001; Triandis, 1995; Uskul & Oyserman, 2005).

Som i monosurveysk sammenhæng peges på, hvad ”socialt ønsket” redigering er, fx af Johnson og Van de Vijver (op.cit.): ”We define social desirability in the response-style view as the tendency of individuals to “manage” social interactions by projecting favourable images of themselves, thereby maximizing conformity to others and minimizing the danger of receiving negative evaluations from them” (p. 196). At blive fri for andres negative eller kritiske vurderinger forvrænger svar og bidrager til at skabe ækvivalensproblemer, fordi respondenter fremstiller sig selv i et mere positivt lys, end hvis de valgte at være sandfærdige (se fx Johnson et al., 2001).

At ”socialt ønsket” redigering af genkaldte informationer kan føre til ækvivalensproblemer fremgår af flere undersøgelser. I et litteraturstudie sammenfattes den hidtil eksisterende viden om, hvordan kulturelle tilhørsforhold påvirker svarredigering (Middleton & Jones, 2000).<sup>54</sup> Det konkluderes, at redigering udgør et problem, der om muligt bør tages højde for ved gennemførelse af multisurveys: ”... individuals are responding, at least in part, with a socially desirable response set consistent with the cultural dimensions predominant in their country or origin”

---

54. Om “socialt ønskede” svar: “... socially desirable response set is the tendency for an individual to deny socially undesirable traits through impressing management and the calculated presentation of a false front, and to admit to socially desirable ones in order to present him/herself in a favourable light” (Middleton & Jones, 2000:151).

(op.cit.:157). Det gælder fx i den tidligere nævnte undersøgelse af Johnson et al. (1997): "Representatives of the cultural groups examined varied significantly in their expression of beliefs and reporting styles thought to be associated with response editing" (p. 107). Ikke kun kulturelle tilhørsforhold, men også indkomst og andre levekår spiller en rolle for, hvordan genkaldte informationer måtte blive redigeret. Alt i alt synes der således at være afgørende forskelle med hensyn til redigering over grænser og på tværs af kulturer: "Persons coming from more influential groups in society or from more affluent countries tend to show lower scores on social desirability" (Johnson & Van de Vijver, i: Harkness et al., 2003:200).

Kulturelt betingede variationer i afgivelse af "socialt ønskede" svar er dog langt fra undersøgt fyldestgørende, hvilket bl.a. fastslås af Johnson og Van de Vijver (i: Harkness et al., 2003; se fx også Blair & Piccinino, i: Hoffmeyer-Zlotnik & Harkness, 2005). Men en del forskningsresultater tyder på, at "socialt ønsket" redigering er mere udbredt i kollektivistiske end i individualistiske lande og kulturer. Om en given kultur er kollektivistisk eller individualistisk påvirker således ikke kun forståelse af spørgsmål og genkaldelse af informationer, men også redigering af sidstnævnte (se fx Lalwani et al., 2006; Smith, i: Presser et al., 2004; Triandis, 1995; Uskul & Oyserman, 2005). Ifølge Triandis (op.cit.) er det fx mere nærliggende at afgive misvisende informationer i kollektivistiske lande og kulturer end i individualistiske, hvorfor også besvarelse af surveyspørgsmål kan blive vildledende. Med en anden formulering af Stocké (2001) er kollektivistiske respondenter bedre til at "gætte", hvad der er "socialt ønskeligt".

Redigeringsforskelle mellem individualistiske og kollektivistiske respondenter fremgår fx af Lalwani et al.'s (op.cit.) undersøgelse blandt amerikanske studerende, der sammenlignes med studerende fra Singapore. Hvor førstnævnte – som tilhørende en individualistisk kultur – ønsker at være autonome og unikke, tilstræber sidstnævnte som kollektivister i højere grad sociale tilhørsforhold og fælles mål. Hvis kollektivismen i højere grad er forbundet med uærlighed og bevarelse af "eget ansigt", mens individualister er mere autentiske, bør de amerikanske respondenter følgelig redigere mindre end Singapore-respondenter, hvilket er hvad Lalwani et al. viser: "... people with both types of cultural orientations or backgrounds engage in desirable responding, albeit in distinct ways" (op.cit.:165).

Hertil kommer, at ”sensitive spørgsmål” er mere sensitive i nogle lande end andre, fx spørgsmål om alkoholforbrug i muslimske lande eller om holdninger til fri abort i lande, hvor svangerskabsafbrydelse ikke er tilladt. Hvis holdninger til fri abort søges afdækket på tværs af lande, vil ækvivalensproblemer bl.a. kunne skyldes, at afgivne svar er redigeret ud fra, om abort er tilladt eller ej: ”The relevance and potentially disturbing role of social desirability is compounded when the different cultures studied have different views on the acceptability of particular behaviors such as abortion, and condom or drug use” (Johnson & Van de Vijver, i: Harkness et al., 2003:202). Desuden er der også forskelle mellem, hvad man efter gældende nationale regelsæt må stille spørgsmål om, fx i Den kinesiske Folkerepublik sammenlignet med USA. Ifølge Smith (i: Presser et al., 2004) må der fx ikke i Kina stilles spørgsmål om KKP (Kinas Kommunistiske Parti).

Til reduktion af svarredigering er en ofte nævnt anbefaling at formulere så lidt ”truende” spørgsmål som muligt: ”... survey designers can both frame questions in a less threatening manner and train interviewers to be nonjudgmental in asking items and reacting to answers” (Smith, i: Harkness et al., 2003:81). Alene mistanke om, at nogle kulturelle grupper kan opfatte et spørgsmål som ”truende”, mens andre måske ikke gør det, bør være tilstrækkeligt til at reformulere eller ekskludere det (se fx Baumgartner & Steenkamp, 2001). Multisurveyforskere bør derfor også være opmærksomme på, at afgørelse af, om et spørgsmål er ”truende” eller ej, fordrer kulturel forståelse og indsigt. Endelig nævnes også tilsikring af anonymitet som et afgørende vilkår for forebyggelse af ækvivalensproblemer.



# AFPRØVNING AF SPØRGESKEMAER

## 5.1. INDLEDNING

Der er i de to foregående kapitler peget på adskillige erfarings- eller forskningsbaserede anbefalinger vedrørende kvalitetssikring af surveydata. Der findes imidlertid yderligere procedurer til kvalitetssikring:

- Hvordan afprøves spørgeskemaer til forebyggelse af multisurveyske måleproblemer?

Blandt procedurer findes de i *oversigt 5.1.* nævnte, hvoraf én velkendt er prøveundersøgelser (afsnit 5.2.). Hertil kommer fokusgruppediskussioner, fx med spørgsmålspræciserende sigte (afsnit 5.3.), og kognitive laboratorieinterview, fx sigtende mod at påpege problemer vedrørende spørgsmålsforståelse (afsnit 5.4.). Yderligere procedurer er interaktionskodning, dvs. kodning af interaktion respondenter og interviewere imellem (afsnit 5.5.), og *split sample*-eksperimenter, hvor stikprøver opdeles i segmenter, der udsættes for forskellige påvirkninger, fx moderat ændrede spørgsmålsversioner (afsnit 5.6.). Endelig skal nævnes debriefing af respondenter, interviewermonitorering, debriefing af interviewere, ekspertpaneler og statistisk ex post-kontrol af data (afsnit 5.7.). En problematiserende sammenfatning findes i rapportens afsluttende kapitel (afsnit 6.2.).

## Oversigt 5.1.

### Procedurer til kvalitetssikring af surveydata.

---

- Kognitive aflastninger
  - Prøveundersøgelser
  - Fokusgrupper
  - Kognitive laboratorieinterview
  - Interaktionskodning
  - *Split sample*-eksperimenter
  - Andre procedurer
- 

Som tidligere sættes der først summarisk fokus på monosurveysk litteratur, hvorpå opmærksomheden rettes imod multisurveysk for hver procedure. Uanset om surveydata overskrider grænser/kulturer eller ej, forudsætter kvalitetssikring et konceptuelt grundlag herfor, eftersom ingen kan frigøre sig fra ontologiske og epistemologiske positioneringer, og fordi ”kvalitet” er et vanskeligt definer- og operationaliserbart begreb. Hvor monosurveyske forfattere som nævnt peger på begreberne *validitet* og *reliabilitet* som bærende kvalitetsbegreber, er det dominerende multisurveyske begreb ækvivalens (se kapitel 2). Trods monosurveysk konsensus om de to første begreber og overvejende multisurveysk enighed om det sidste, er der imidlertid ikke konsensus om, hvordan begreberne bør defineres alment.

Implementering af de omtalte procedurer skal således bidrage til at sikre monosurveyske datas validitet og reliabilitet, dvs. til i videst muligt omfang at reducere, hvad der tidligere er fællesbetegnet som måleproblemer. Til virkeliggørelse heraf peger adskillige forskere på, at reduktion af måleproblemer kun er mulig, hvis procedurer iværksættes med afsæt i tydeligt formulerede formål: “A number of researchers have pointed out that there are many purposes for pretesting and that investigators will maximize their results by specifying clear and precise pretest objectives” (Czaja, 1998; se fx også Converse & Presser, 1986).

Når det derimod handler om kvalitetssikring af multisurveydata, skal forskellige procedurer tilsikre en ækvivalens, der kun vagt eller slet ikke er alment defineret, og som findes i over 50 forskellige former (se kapitel 2). Nogle multisurveyske forfattere pointerer, at én eller flere former for prætest af spørgeskemaer er særlig vigtige, når grænser eller kulturer overskrides (se fx Hudler & Richter, 2002; Smith, i: Presser et al., 2004a). Navnlige fordi oversættelser kan generere yderligere ækvivalensproblemer, er det ikke tilstrækkeligt at afprøve et spørgeskema i én

national eller kulturel kontekst (se fx Behling & Law, 2000; Birbili, 2000). Det nævnes ligeledes, at afprøvning af spørgeskemaer forudsætter samarbejde. Eksempelvis mener Pasick et al. (1996), at prætest bør være "... a team effort with multiple disciplines and preferably multiple cultures represented" (p. 530).

Trods anbefalinger af flere af de i kapitlet nævnte procedurer, synes ingen multisurveyiske anbefalinger at være baseret på systematisk empirisk forskning: "Two major obstacles to effective developmental work in cross-national surveys are the absence of methodological studies of the various pretesting approaches and a general underutilization of pretesting" (Smith, i: Presser et al., 2004a:450). Ingen undersøgelser sætter med andre ord tværnationalt eller -kulturelt fokus på systematiske forskelle og ligheder mellem forskellige former for afprøvning af multisurveyiske spørgeskemaer. Når undersøgelser undertiden belyser problematikken, er det inden for rammerne af nationale optikker (se fx Blair & Piccinino, i: Hoffmeyer-Zlotnik & Harkness, 2005; Presser & Blair, 1994).

## 5.2. PRØVEUNDERSØGELSER

Mange monosurveyiske forfattere anbefaler prøveinterview (se fx Aldridge & Levine, 2001; Alreck & Settle, 1995; Czaja & Blair, 1996; de Vaus, 1998; Fink, 1998; Fowler, 1993; Fowler & Mangione, 1990; Krosnick, 1999; Presser et al., 2004; Willis et al., 1991). Hermed konfronteres spørgeskemaer så at sige med virkelighedens verden. Prøveundersøgelser belyser dermed, hvordan spørgsmål og spørgeskemaer fungerer "i praksis": "Once a questionnaire has been developed, each question and the questionnaire as a whole must be evaluated rigorously before final administration" (de Vaus, 1998:99). Ofte påpeges dog svagheder ved prøveinterview, hvorfor disse af mange metodeforfattere foreslås suppleret med andre procedurer.

Men hvad er en prøveundersøgelse i monosurveyisk optik? Aldridge & Levine (2001) definerer fx prøveundersøgelser således: "A pilot survey is a dummy run of the survey proper, in which we aim to test all the key aspects of the survey" (Aldridge & Levine, 2001:90), mens Czaja & Blair (1996) har en lidt anden opfattelse: "In a conventional pretest, a small number of respondents ... are interviewed. The exact

procedures planned for the main study are followed. After this interviewing takes place, a debriefing meeting is held with the interviewers” (Czaja & Blair, 1996:95). Hermed antydes, at der ikke er konsensus om krav til prøveundersøgelser. Én ledetråd er dog, at prøveundersøgelser bør gennemføres på samme måde som i egentlige monosurveys, dvs. under realistiske betingelser.

Hvad er prøveundersøgelsers fordele ifølge monosurveysk litteratur? Til fordelene hører, at de som regel giver anledning til at påpege svagheder ved spørgeskemaer. Prøveundersøgelser er dog ingen garant for reduktion af måleproblemer, men bedre end ingenting. Fowler & Mangione (1990) peger fx på, at ikke alle problemer lader sig forudsige, hvorfor besøg ”i marken” er nødvendige: “... some questions pose virtually impossible tasks for respondents and interviewers” (Fowler & Mangione, 1990:94). Hertil kommer, at prøveundersøgelser kan belyse, om visse spørgsmål stiller for store kognitive krav (se fx Fowler & Cannell, i: Schwarz & Sudman, 1996).

Der er imidlertid også kritiske røster: “... there is no scientific evidence to justify the confidence that this kind of pretesting identifies the major problems in a questionnaire” (Presser et al., 2004). Prøveinterview er ofte for få, for usystematiske og for overfladiske (Se fx Aldridge & Levine, 2001; Fowler & Cannell, i: Schwarz & Sudman, 1996; Krosnick, 1999; Presser et al., 2004). Utilstrækkelige økonomiske midler nævnes ofte som begrundelse for ikke at gennemføre kvalitetssikring af surveydata. Ydermere tilbyder prøveundersøgelser begrænset indsigt i respondenters problemer med at forstå, genkalde informationer i forhold til og besvare spørgsmål (se fx Fowler, i: Bickman & Rog, 1998). Undersøgelser af ældre dato underbygger kritikken, da det har vist sig, at prøveundersøgelser fx er uegnede til at afdække forskellig forståelse af enkeltord og spørgsmål (se fx Belson, 1986).

Også multisurveysk litteratur sætter fokus på prøveundersøgelser, der efter enkeltes opfattelse kan bidrage til reduktion af ækvivalensproblemer, men almindeligvis menes andre procedurer at være mere anbefalelsesværdige, fx kognitive interview og ekspertvurderinger (se fx Blair & Piccinino, i: Hoffmeyer-Zlotnik & Harkness, 2005; Birbili, 2000; Hudler & Richter, 2002; Smith, i: Presser et al., 2004a). Ligesom en del monosurveyske forfattere påpeger også multisurveyske konventionelle prøveundersøgelsers ulemper, men på den anden side er den slags prætest bedre end intet at gøre. Undertiden understreges desuden, at prøve-



undersøgelser er endnu vigtigere ved gennemførelse af multi- end monosurveys.

### 5.3. FOKUSGRUPPER

Yderligere en procedure til sikring af datakvalitet er fokusgrupper. Proceduren, der normalt benyttes i datatilvirkningens indledende fase, anbefales i dele af monosurvey-sk metodelitteratur og synes at finde voksende udbredelse, fx i USA (se fx Alreck & Settle, 1995; Czaja, 1998; Fowler, 1995; Fowler & Cannell, i: Schwarz & Sudman, 1996; Nassar-McMillan & Borders, 2002; Rea & Parker, 1997). Når monosurveyforskere advokerer for fokusgrupper, er det bl.a. under henvisning til, at gruppemedlemmernes interaktion tilbyder forskere informationer, som næppe ville kunne opnås på andre måder: "Focus groups involve the explicit use of group interaction to produce data and insights that would be less accessible without the interaction found in a group" (Nassar-McMillan & Borders, 2002).

Ved gennemførelse af fokusgruppediskussioner udvælges og mødes et begrænset antal respondenter for at drøfte på forhånd fastlagte såvel som situationelt opståede problematikker (se fx Alreck & Settle, 1995). Fokusgrupper kan sigte imod at præcisere problemstillinger, men kan også benyttes ved afprøvning af spørgsmåls relevans: "Before writing a draft of a structured set of questions, it is almost always valuable to conduct focused discussions with people who are in the study population about the issues to be studied" (Fowler, 1993:95). Ifølge Fowler forudsættes, at forskere har nogenlunde præcise forestillinger om, hvordan spørgeskemaer ender med at blive (Fowler, i: Bickman & Rog, 1998).

Også spørgeskemaudkast kan afprøves i grupper, fx med hensyn til forståelse af spørgsmål og genkaldelse: Påfører forskere respondenter opgaver større, end de kan løse? Forstås spørgsmål som tilsigtet? Har respondenter mulighed for at genkalde informationer? Fokusgruppediskussioner kan også belyse, om respondenter er villige til at afgive nøjagtige svar: "Focus groups are very good for determining respondents' understanding of key terms and concepts; how respondents recall or retrieve information; whether behavioral frequencies are being

counted, estimated, or "calculated" using some other strategy" (Czaja, 1998).

Trods adskillige monosurveyske metodeforfatteres anbefaling af fokusgrupper er litteraturen kendetegnet ved manglende konsensus om, hvordan fokusgruppediskussioner gennemføres (se fx Nassar-McMillan & Borders, 2002). Mens nogle peger på, at deltagerantallet ikke bør overskride 6-8 gruppemedlemmer, mener andre, at antallet kan variere fra gruppe til gruppe. Hvordan gruppemedlemmer udvælges, er der heller ikke enighed om. Hvor nogle forfattere anbefaler heterogenitet, peger andre på, at der kan argumenteres for homogen udvælgelse: "... homogeneity or heterogeneity of group members depends on whether any of the topics will make participants feel uncomfortable or reluctant to discuss their thoughts" (Czaja, 1998). Flere forfattere pointerer, at forskere, der gennemfører fokusgruppediskussioner, bør have erfaring hermed (se fx Nassar-McMillan & Borders, 2002).

Til procedurans fordele hører bl.a., at den kendetegnes ved fleksibilitet og detaljerede svar, at omfattende informationsmængder kan genereres på kort tid, at fremgangsmåden fører til uforudsete oplysninger, og at anvendte resultater forbedrer spørgeskemaer (se fx Czaja, 1998). Som ulemper påpeges, at fokusresultater ikke kan overskride deltageres ytringer, dvs. ikke lader sig generalisere, og at udbyttet undertiden er mangelfuldt, fordi mange synes uden erfaring med gennemførelse af fokusgruppeinterview. Det nævnes også, at enkelte respondenters gruppedominans og afvisning af sensitive emner er mulige dysfunktioner.

I multisurveysk litteratur nævnes og anbefales fokusgrupper i begrænset omfang (se fx Blair & Piccinino, i: Hoffmeyer-Zlotnik & Harkness, 2005; Harkness, i: Harkness et al., 2003; Harkness et al., i: Presser et al., 2004a; Singh et al., 2000). Det kan bl.a. skyldes, at fokusgruppedrøftelser ofte finder sted til præcisering af problemstillinger og i datatilvirkningens indledende fase, mens tværnationale surveyundersøgelser ofte gennemføres ved at "overføre" et masterspørgeskema fra ét land og ét sprog til andre. Da selv "gode" oversættelser ikke nødvendigvis "fungerer" i praksis, er afprøvning heraf afgørende, påpeger Harkness et al. (i: Presser et al., 2004a) og andre, der samtidig er af den opfattelse, at fokusgrupper kan være velegnede til fx at afprøve spørgeskemaer oversat til andre sprog. I litteraturen nævnes da også enkelte eksempler på anvendelse af fokusgrupper, eksempelvis test af semantisk ækvivalens i form af båndoptagede fokusgruppeinterview med afro-amerikanere og

kaukasere: "... cultural differences are apparent in the way parents of children with emotional and behavioral disorders interpret some of the items of a commonly used instrument" (Singh et al., 2000:132).

#### 5.4. KOGNITIVE LABORATORIEINTERVIEW

Kognitive laboratorieinterview er nævnt som en procedure, der kan forbygge måleproblemer. Herom findes en righoldig monosurveysk litteratur, der for en dels vedkommende betoner prøveundersøgelsers og fokusgrupperes utilstrækkelighed (se fx Aborn, i: Sirken et al., 1999; Czaja, 1998; DeMaio et al., 2003; Forsyth & Lessler, i: Biemer et al., 1991; Fowler, 1995; Fowler, i: Bickman & Rog, 1998; Jobe & Mingay, 1991; Krosnick, 1999; Lessler et al., 1989; Presser et al., 2004; Sirken & Schechter, i: Sirken et al., 1999; Sudman et al., 1996; Tourangeau et al., 2000; Willis et al., 1991; Willis et al., i: Sirken et al., 1999). Laboratorieinterview, der drager nytte af lingvistik og kognitionspsykologi, "... are increasingly seen as an essential step in the design and evaluation of a survey instrument. Questions that are not consistently understood or answered in a laboratory setting certainly will not work any better in an actual survey" (Fowler, 1993:98).

Hvad er kognitive laboratorieinterview i monosurveysk optik? Som andre spørgsmål besvares heller ikke dette konsensuelt, men der er enighed om at anskue den slags interview som verbale "hjernebesøg", hvis formål er at belyse kognitive svarprocessuelle aspekter (se fx Czaja, 1998). Czaja & Blair (1996) tildeler fx kognitive interview følgende kendetegn: "Borrowing a procedure called think aloud from cognitive psychology, respondents are instructed to think out loud as they answer each question. The idea is to determine what things respondents consider in answering questions" (Czaja & Blair, 1996:99). En lidt anderledes opfattelse er Fowler's (1998), der peger på tre træk: Interview sigter imod at belyse spørgsmålsforståelse og andre svarprocessuelle aspekter, de foregår normalt under laboratorieforhold, og ved at "tænke højt" bidrager respondenter til forståelse af svarprocessen.

Når det handler om gennemførelse af laboratorieinterview, er monosurveyske svar forskellige. På linje med kvalitativ metodelitteratur er der fx ikke konsensus om, hvordan respondenter, der skal interviewes, bør selekteres, ligesom også anbefalinger vedrørende antallet heraf er

forskellige. Én besvarelse er DeMaio's (2003): "The number of interviews is small (typically 20 or less). The respondents are not scientifically selected; rather, subjects are purposively recruited to permit testing of separate parts of the questionnaire (...) or the same part of the questionnaire by different types of people" (DeMaio, 2003). Andre forfattere anbefaler, at respondenter selekteres efter heterogenitetskriterier (se fx Fowler, 1995).

I flere monosurveyske fremstillinger nævnes, at laboratorie-interview, der normalt båndoptages og analyseres, gennemføres "under fire øjne", men der er ikke enighed om, hvem der bør virke som interviewere. Nogle pointerer, at laboratorieinterview fordrer omfattende kvalifikationer og derfor bør gennemføres af forskere med kognitionspsykologisk ekspertise (se fx Fowler, 1995). Der er heller ikke konsensus om, *hvor* interview anbefales gennemført: "Cognitive interviews are considered a 'laboratory' method ... and they typically take place in a laboratory, although they can also take place in the respondent's home or in a central location ..." (DeMaio et al., 2003). Mens nogle forskere insisterer på, at kognitive interview altid bør foregå under laboratorielignende forhold, påpeger andre, at de også lader sig gennemføre i "feltet" (se fx DeMaio & Rothgeb, i: Schwarz & Sudman, 1996).

Laboratorieinterview er tænke-højt-interview, der findes i to versioner: retrospektive eller kontinuerlige. Når interview er kontinuerlige, tænker interviewpersoner højt, mens spørgsmål besvares, mens retrospektive interview forudsætter ex post tænken højt, dvs. efter afsluttet gængs interviewing. I "genuine" kontinuerlige interview er intervieweren næsten tavs og intervenserer kun, hvis højt-tænkning er utilstrækkelig (se fx DeMaio et al., 2003). I interviewpraksis findes imidlertid yderligere aspekter. Det drejer sig bl.a. om tænke-højt-fremmede brug af *probes* (se fx Czaja, 1998).

Ifølge monosurveysk litteratur gennemføres tænke-højt-interview sjældent i "ren" form, men suppleres med *probes*, parafrasering, *confidence ratings*, *response latency*, *memory cues* og genbesvarelse af spørgsmål (se fx Fowler, 1993; Jobe & Mingay, 1989). Spørgsmålsparafrasering er én måde at belyse forståelse på. Det er dog omdiskuteret, om parafrasering afspejler forståelse, fx fordi nogle respondenter har lettere ved at parafasere end andre (se fx Foddy, 1993). Et andet aspekt er *confidence ratings*, dvs. at spørge, hvor sikker respondenter er på svars nøjagtighed (se fx Jobe & Mingay, 1989). Også tidsanvendelse fra et spørgsmål stilles,

til svar påbegyndes, kan være en vigtig problemindikator (se fx Czaja, 1998). Tilsvarende gælder *memory cues*, dvs. afprøvning af, hvordan tidsmæssige og andre ”hukommelsesankre” fungerer. Endelig kan også ændret genbesvarelse af tidligere i retrospektive laboratorieinterview besvarede survey spørgsmål virke som indikator på måleproblemer (se fx Fowler, 1995).

Sammenfattende synes laboratorieinterview at have betydelige fordele ifølge monosurvey litteratur. Ved velforberejdede ”hjernebesøg” tilvirkes informationer, som fx prøveundersøgelser ikke tilbyder: “... the method is very useful for providing input about how respondents actually formulate their answers and the kinds of errors they introduce that is not available through any other method” (DeMaio et al., 2003). Hertil fordres dog, at forskere, der gennemfører laboratorieinterview, har kognitionspsykologisk ekspertise og erfaring med gennemførelse af kvalitative interview.

Der er også advarende monosurvey røster, fx at laboratorieinterview adskiller sig fra gængs *common ground* respondenter og interviewere imellem, og at ikke alle er lige egnede til at verbalisere (se fx Czaja, 1998). Andre påpeger laboratorieinterviews manglende generaliseringsmuligheder som en svaghed. Enkelte er skeptiske over for laboratorieinterview og spørger, hvad fx amerikanske surveyforskere generaliserende har lært om spørgeskemakonstruktion ved gennemførelse af tusinder af laboratorieinterview (se fx Willis et al., i: Sirken et al., 1999).

Kognitive laboratorieinterview nævnes og anbefales ofte i multisurvey litteratur (se fx Blair & Piccinino, i: Hoffmeyer-Zlotnik & Harkness, 2005; Carrasco, 2003; Forsyth, 2006; Goerman, 2006; Harkness, i: Harkness et al., 2003; Jobe & Mingay, 1990; Johnson, 1998; Miller, 2003; Miller, 2004; Miller et al., i: Hoffmeyer-Zlotnik & Harkness, 2005; Smith, i: Presser et al., 2004a; Warnecke et al., 1997; Willis et al., 2005; Willis, i: Cohen & Lepkowski, 2004). Ifølge flere forskere er tænkehøjt-interview velegnede til belysning af, hvordan kulturelt forskellige respondenter løser semantiske og andre kognitive opgaver. Det gælder fx Blair og Piccinino (op.cit.): “The main strength of cognitive interviewing is its potential ability to identify problems at any stage of the response process” (p. 22). Samme opfattelse har Miller et al. (op.cit.), der anbefaler kvantificerende kodning: “... various types of analyses can be conducted using coded, semi-structured cognitive interviews which examine relationships between participant characteristics, such as ethnicity, and re-

sponses to tested survey questions. Further, based on quantification of these codes, it may be determined that certain problems are more serious than others” (p. 89).

Laboratorieinterviews egnethed fremhæves endvidere i forbindelse med afprøvning af spørgeskemaer oversat til andre sprog, fx af Forsyth (2006): “When cognitive interviews are used to test a questionnaire translation, results can be used to identify additional translation deficiencies. For example, target-language terms that respondents interpret differently than intended, target-language terms that are unfamiliar, or target-language terms that have culture-specific meanings”. Som i monosurveysk litteratur er der imidlertid også kritiske røster, fx vedrørende praktisk planlægning og gennemførelse af kognitive interview (se fx Willis, i: Cohen & Lepkowski, 2004). Eksempelvis skal relevante informanter selekteres, resultater kan ikke generaliseres, resultater kan være mangelfulde, og der kan være problemer med at finde kvalificerede interviewere.

Trods manges anbefalinger er kvalitetssikrende kognitive interview og analyser heraf dog endnu ret sjældne i multisurveysk sammenhæng, men svarende til den monosurveyske udvikling forventer Miller et al. (i: Hoffmeyer-Zlotnik & Harkness, 2005) og andre, at proceduren efterhånden finder voksende multisurveysk udbredelse. Nogle af de ovenfor nævnte litteraturhenvisninger angår anvendelse af kognitive interview i konkrete multisurveys (se fx Carrasco, 2003; Forsyth, 2006; Miller, 2003; Miller, 2004; Warnecke et al., 1997; Willis et al., 2005).

Ét eksempel omhandler kognitive interview med kulturelt forskellige respondenter i Mississippi-området og sætter fokus på kulturelle tilhørsforholds svarprocessuelle betydning (Miller, 2003; Miller, 2004). Undersøgelsen viser, at spørgeskemaer bør – og ved gennemførelse af kognitive interview kan – tilpasses forskellige gruppers kulturelle tilhørsforhold: “... cognitive analysis of interviews with different cultural and racial/ethnic groups suggests several guidelines for improving questionnaires so that survey research can advance the quality and estimates for the populations” (Miller, 2003:188). Når respondenter fx aldrig tidligere har deltaget i surveyundersøgelser, er det vanskeligt for dem at forstå og honorere den af forskere forventede spørgsmål/svar-proces.

I yderligere et eksempel undersøger Warnecke et al. (1997) tværkulturelle forståelser af, hvordan *fysiske aktivitet* opfattes: “Variation by race/ethnicity was found in respondent interpretation of what consti-

tutes physical activity through probes asking whether they considered walking, household, work-related activity, and yard work to be physical activities” (p. 337). Et sidste eksempel handler om udfordringer i forbindelse med oversættelse af spørgeskemaer og kvalitetssikrende anvendelse af kognitive interview: ”Translation difficulty should not be underestimated as a challenge to cross-cultural equivalence, even if ‘best practices’ of translation have been used” (Willis et al., 2005:8). Ifølge Willis et al. (op.cit.) viser erfaringer fra tre tværkulturelle helbredssurveys, at kognitive interview er velegnede til afprøvning og tilpasning af oversættelser, fx fra engelsk til spansk.

## 5.5. INTERAKTIONSKODNING

Mens kognitive laboratorieinterview søger at opnå adgang til respondenteres svarprocesser, er formålet med interaktionskodning systematisk registrering af samspillet mellem respondenter og interviewere for ad den vej at reducere måleproblemer. Herom findes en righoldig monosurveysk metodelitteratur, hvori beskrives, hvad interaktionskodning er, hvordan den gennemføres og hvilke fordele og ulemper, der knytter sig til proceduren (se fx Czaja, 1998; DeMaio et al., 2003; Esposito et al., 1991; Fowler, i: Bickman & Rog, 1998; Fowler & Cannell, i: Schwarz & Sudman, 1996; Fowler, 1993; Fowler, 1995; Oksenberg et al., 1991; Peterson, 2000; Sudman et al., 1996).

Dele af monosurveysk metodelitteratur peger på, at prøveundersøgelser tilbyder utilstrækkelige muligheder for forebyggelse af måleproblemer. Da båndoptagelse og kodning af interaktioner respondenter og interviewere imellem viser, hvordan spørgeskemaudkast fungerer, er interaktionskodning et vigtigt supplement. Interaktionskodning (*behavior coding*) er en procedure udviklet af Charles Cannell m.fl. ved University of Michigan (se fx Fowler & Cannell, i: Schwarz & Sudman, 1996). Til at begynde med sigtede proceduren kun imod overvågning af intervieweradfærd, men blev udviklet til at belyse interaktion respondenter og interviewere imellem for derigennem bl.a. at forbedre spørgeskemaer (se fx DeMaio, 2003).

Nogle monosurveyske forfattere mener, at proceduren er nogenlunde ”objektiv”: ”The coding of interviewer-responder exchanges represents a relatively objective method of identifying questionnaire

items that are causing problems for interviewers and respondents” (Esposito et al., 1991:48). Ikke desto mindre indskrives den sig i en særlig interviewopfattelse. Det er ikke den eneste præmis, da proceduren baseres på “... the assumption that any deviation from the ideal model, in which an interviewer reads a question exactly as written and the respondent provides a complete answer, indicates a problem with the question” (Czaja, 1998). Ifølge Fowler er rationalet, der matcher en stimulus-respons-opfattelse af surveyinterview: “When a survey interview is going perfectly, the interviewer will read the question exactly as written once, after which the respondent will give an answer that meets the question objectives. When there are deviations from this, the perfect question-and-answer process, it may be an indication of a question problem” (Fowler, i: Bickman & Rog, 1998:367).

Proceduren sigter altså imod at registrere ”afvigelser” fra en ”ideel”, men normativ model: “In interaction coding, a third party uses a simple set of codes to indicate ... nonverbatim reading of the question, respondents’ requests for clarification, or inadequate answers” (Czaja & Blair, 1996:101). “Afvigelser” kan også være afbrydelse af spørgsmålsoplæsning, afgivelse af ved ikke-svar og svarnægtelse (se fx Sudman et al., 1996). I metodelitteraturen findes adskillige eksempler på kodeskemaer, der lader sig anvende ubesværet (se fx Fowler, 1995). Nogle metodeforfattere peger på, at andelen af ”afvigelser” ikke bør overstige et på forhånd berammet niveau, fx 15 pct. (se fx Oksenberg et al., 1991). Er andelen højere, bør det give anledning til revision af relevante dele af spørgeskemaet.

Især metodeforfattere, der bifalder stimulus-respons-modellen, mener, at interaktionskodning er anvendelig og samtidig en enkel, fleksibel og billig procedure til forebyggelse af måleproblemer (se fx Fowler, 1995). Hvad fleksibilitet angår, påpeger DeMaio (2003): “The method is flexible in that the coding scheme used in any particular application can be adapted to meet specific priorities of the questionnaire designers” (DeMaio et al., 2003). Det pointerer endvidere, at interaktionskodning er anvendelig, fordi den tilføjer viden om, hvordan spørgsmål fungerer, og at kodningen derfor kan yde væsentlige bidrag til forbedring af spørgeskemaer. Som yderligere fordele nævnes objektivitet, gentagelighed og sammenlignelighed (se fx Fowler & Cannell, i: Schwarz & Sudman, 1996:17).



Der er dog også kritiske monosurveyske røster, hvoraf én angår interaktionsregistrering, der kan være behæftet med fejl, og som aldrig er ”objektiv” (se fx Sykes & Morton-Williams, 1987). Det er fx ikke enkelt at afgøre, om datatilvirkning er i overensstemmelse med gældende interviewretningslinjer. Et mere afgørende aber dabej er dog, at interaktionskodning ikke belyser afvigelsers årsager (se fx Czaja, 1998). I så henseende påpeger Houtkoop-Steenstra (2000), at konversationsanalyser kan være nyttige, fordi de bidrager til at forklare, hvorfor interaktionsproblemer opstår.

I multisurveysk litteratur nævnes interaktionskodning som en mulig procedure til reduktion af ækvivalensproblemer (se fx Blair & Piccinino, i: Hoffmeyer-Zlotnik & Harkness, 2005; Johnson, 1998; Smith, i: Harkness et al., 2003; Van der Zouwen & Smith, 2004; Willis, i: Cohen & Lepkowski, 2004). Willis (op.cit.) gør fx opmærksom på, at interaktionskodning i højere grad er en kvantitativ end kvalitativ procedure, og at den ”... involves the analysis of interactions between the interviewer and respondent in search of overt indications that difficulties may exist” (p. 95). At tegn på eventuelle interaktionsproblemer skal være åbenbare, er et problem, fordi problemer kan være til stede, uden at de kommer til syne.

Sammenlignet med tilslutningen til laboratorieinterview anbefaler kun få forfattere fremgangsmåden, formodentlig bl.a. fordi den som nævnt indskriver sig i en stimulus-respons-interviewopfattelse, der er særlig vanskelig at praktisere, når grænser eller kulturer overskrides. Blair & Piccinino (op.cit.) problematiserer fx, om samme adfærdskoder er forventelige på tværs af grænser eller kulturer, og om respondenter, der deltager i multisurveys, afgiver svar på måder, der er sammenlignelige. Som følge heraf indeholder litteraturen kun få eksempler på procedurens praktiske anvendelse i multisurveysk sammenhæng (se fx Johnson, 1998).

## 5.6. *SPLIT SAMPLE*-EKSPERIMENTER

Spørgeskemaer kan også afprøves ved gennemførelse af *split sample*-eksperimenter, der ofte omtales i monosurveysk metodelitteratur (se fx Aldridge & Levine, 2001; Fowler, 1993; Groves, 1989; Krosnick & Alwin, 1987; Litwin, 1995; Smith, i: Biemer et al., 1991; Schuman et al., 1983; Schuman & Presser, 1996; Smith, i: Biemer et al., 1991; Touran-

geau & Rasinski, 1988). Mens resultater af prøveundersøgelser, fokusgrupper, laboratorieinterview og – i al fald efter nogle metodeforfatteres opfattelse – interaktionskodning ikke formår at tilvejebringe generaliserende viden om måleproblemers udbredelse, gælder det modsatte *split sample*-eksperimenter.

*Split sample*-procedurer indebærer, at stikprøver dekomponeres i to eller flere segmenter, hvoraf det ene interviewes eller besvarer postspørgeskemaer med afsæt i én version af et spørgeskema, mens det andet segment besvarer spørgsmål tilhørende en anden version (se fx Aldridge & Levine, 2001). Kun fantasien sætter grænser for, hvordan der i to spørgeskemaversjoner fx kan manipuleres med syntaks, valg af ord, transformation af spørgsmål, ændrede svarkategorier, spørgsmålsrækkefølge og genkaldelsesfremmende vink. Efterfølgende kan der drages statistiske slutninger om forskelle to eller flere versioner imellem.

*Split sample*-eksperimenter gennemføres ofte til belysning af forståelse af ord, spørgsmålskategorier, svarkategorier, spørgsmålsrækkefølge mv. Adskillige af de i afsnit 4.2. nævnte eksempler på forståelser af moderat ændrede spørgsmål baseres på *split sample*-data. Eksperimenter af den type har bidraget til voksende monosurveysk erkendelse af, hvor sensitive respondenter er over for spørgsmåls formulering (se fx Schuman & Presser, 1996). Tilsvarende gælder ændrede svarkategorier, fx lukkede spørgsmål ændret til åbne og tilføjede ved ikke-kategorier (se også Olsen, 2001a).

*Split sample*-eksperimenter anvendes også i andre monosurveyske sammenhænge, fx til belysning af måleproblemer grundet genkaldelsesproblemer (se fx Blair & Burton, 1987). Endelig er der også gennemført adskillige *split sample*-eksperimenter til belysning af konteksteffekter. Tourangeau & Rasinski (1988) ”jonglerede” fx med 19 forskellige versioner af ”samme” spørgeskema og fandt markante konteksteffekter, mens Smith (i: Biemer et al., 1991) ved anvendelse af *split sample*-proceduren kun afdækkede få og moderate konteksteffekter.

*Split sample*-eksperimenters fordel er, at målingers stabilitet eller mangel på samme induktivt-statistisk overskrider respondenteres besvarelser og dermed kan populationsgeneraliseres. Proceduren har dog også en svaghed, nemlig at det normalt er umuligt at pege på den spørgeskemaversion, der skaber færrest måleproblemer. Kun hvis resultater sammenholdes med andre informationer, er lokalisering af det ”bedste” spørgsmål, de ”bedste” svarkategorier eller de mindste konteksteffekter

mulig. Andre informationer kan være valide registeroplysninger, fx om indkomst, formue, boligstandard, lægebesøg, hospitalsindlæggelser eller psykiske lidelser.

I multisurveysk litteratur nævnes *split sample*-eksperimenter kun af få forfattere (se fx Braun & Scott, 1998; Harkness, i: Harkness et al., 2003; Harkness et al., i: Presser et al., 2004a:469). Braun og Scott (op.cit.) anbefaler fx den slags eksperimenter til kontrol af funktionel ækvivalens, mens Harkness (op.cit.) fx peger på, at *split sample*-tilgangen er anvendelig ved afprøvning af spørgeskemaer oversat til andre sprog. At proceduren nævnes af få forfattere og anbefales af endnu færre skyldes næppe, at den principielt skulle være uanvendelig i multisurveys, men nok snarere en række praktiske og tidsmæssige problemer i tilknytning til *split sample*-eksperimenter, der overskrider grænser eller kulturer.

## 5.7. ANDRE PROCEDURER

Med fremstillingen ovenfor er fem procedurer til afprøvning af spørgeskemaer præsenteret og problematiseret. I monosurveysk litteratur nævnes imidlertid yderligere procedurer, der kan bidrage til reduktion af måleproblemer: debriefing, monitorering af interviewere og ekspertvurderinger.

Hvad respondentdebriefing angår, er proceduren – undertiden kaldet *frame of reference probing* – beslægtet med laboratorieinterview: ”In general, the goals of respondent debriefing are the same as for intensive individual or cognitive interviews” (Czaja, 1998). Men i modsætning til laboratorieinterview foregår respondentdebriefing i ”feltet”, almindeligvis i forlængelse af prøveinterview. Som nævnt i afsnit 5.4. insisterer nogle forfattere på, at kognitive interview bør foregå under laboratorielignende forhold, mens andre påpeger, at de også lader sig gennemføre i ”feltet”: ”Our view is that cognitive methods need not be restricted to the laboratory” (DeMaio & Rothgeb, i: Schwarz & Sudman, 1996:177).

Én fordel ved debriefing af respondenter er, at proceduren i højere grad end laboratorieinterview muliggør generaliseringer, påpeger flere forfattere. Det skyldes, at proceduren er struktureret og almindeligvis gennemføres med flere respondenter end ved laboratorieinterview: ”... respondent debriefing is structured and is conducted on a larger scale as part of a field pretest. It is designed to yield results that are more

generalizable than those obtained through laboratory interviews. Because of the large sample sizes used in field tests, respondent debriefing can also provide information about rare situations” (DeMaio & Rothgeb, i: Schwarz & Sudman, 1996:188; se også Esposito et al., 1991). Der er imidlertid også advarende røster, hvoraf én angår mulige efterrationaliseringer.

Monitorering og debriefing af interviewere er andre procedurer, der kan forebygge måleproblemer. Monitorering findes i forskellige former, fx overvågning af interviewøvelser i forbindelse med uddannelse af interviewere. Overvågning af egentlige interview er også en anbefalet procedure (se fx Czaja & Blair, 1996; Lyberg & Kasprzyk, i: Biemer et al., 1991). I forbindelse hermed nævnes, at telefoninterview er lettere at supervisere end besøgsinterview, idet sidstnævnte dog kan optages på bånd og ”superviseres” ex post. Også direkte observation af besøgsintervieweres adfærd er en mulighed (se fx Fowler & Mangione, 1990).

Ligesom debriefing af respondenter er én mulig procedure, gælder tilsvarende debriefing af interviewere (se fx Czaja, 1998; Czaja & Blair, 1996; Fowler, 1993). Flere monosurveyske forfattere nævner eller underforstår interviewerdebrieffing som en del af ”klassiske” prøveundersøgelser. Ved besvarelse af *follow up*-spørgsmål kan interviewere fx bidrage til formulering af ”gode” spørgsmål. Der er dog også advarende røster: “Interviewer debriefing sessions ... can also be misleading. Interviewers typically complete only a handful of interviews for a pretest (Czaja, 1998). Der peges tillige på, at det for mange interviewere kan være problematisk både at virke som interviewere og spørgeskemaevaluatore.

Endelig kan ekspertvurderinger, der ifølge flere monosurveyske forfattere ikke er et substitut fx for laboratorieinterview, bidrage til reduktion af måleproblemer, fx ved at forudsige semantiske og andre kognitive tvivlsspørgsmål (se fx Czaja, 1998; Dillman & Tarnai, i: Biemer et al., 1991; Lessler & Forsyth, i: Schwarz & Sudman, 1996; Presser & Blair, 1994). Der peges på, at ekspertvurderinger af spørgeskemaer er en procedure med stort udbytte og relativt lave omkostninger (se fx Czaja, 1998). Ekspertvurderinger kan være individuelle eller gennemføres som fokusgruppediskussioner, fx baseret på på forhånd besvarede evalueringsskemaer.

Også multisurveysk litteratur sætter fokus på procedurer, der overskrider prøveundersøgelser, fokusgrupper, kognitive laboratoriein-

terview, interaktionskodning og *split sample*-eksperimenter (se fx Behling & Law, 2000; Blair & Piccinino, i: Hoffmeyer-Zlotnik & Harkness, 2005; Harkness & Schoua-Glusberg, 1998; Hines, 1993; Johnson, 1998; Johnson et al., 1997; McKay et al., 1996; Smith, i: Presser et al., 2004a; Willis, i: Cohen & Lepkowski, 2004). Det drejer sig om de i monosurveysk sammenhæng nævnte, hvor der navnlig advokeres for ekspertvurderinger af spørgeskemaer og oversættelser. Flere forfattere pointerer dog som betingelse, at eksperter har specialviden om netop de sprog og kulturer, som en given multisurvey sætter fokus på (se fx Blair & Piccinino, op.cit.).

Som ét eksempel på anvendelse af eksperter kan nævnes etnografer med dybdeforståelse af multisurveysk relevante kulturer (se fx Word, 1992). Etnografer eller andre kvalitative forskere kan yde væsentlige bidrag til forebyggelse af ækvivalensproblemer. Yderligere eksempler er Behling & Laws (2000) anbefaling om inddragelse af eksperter ved gennemførelse af modifieret direkte oversættelse (se afsnit 3.4.) eller McKay et al.'s (1996) forslag om sprogkyndiges udarbejdelse af oversættelsesreviews. Mens nogle forfattere anbefaler brug af eksperter ved afprøvning af spørgeskemaer, opfordrer andre til at konsultere eller samarbejde med eksperter i hele undersøgelsesforløbet, fx for at forebygge etnocentriske tendenser (se fx Johnson, 1998; Van de Vijver & Hamblen, 1996).<sup>55</sup>

---

55. Da rapportens tema er konstruktion af multisurveydata og kvalitetssikring heraf, overskrider statistiske ex post-afprøvninger af data temaet. Det findes forskellige typer af statistiske procedurer, som anbefales og anvendes af mange multisurveyforskere, fx faktoranalyse (se fx Blasius & Thiessen, 2006; Braun & Scott, 1998; King et al., 2003; Moors, 2004; Mullen, 1995; Patel et al., 2005; Pauwels & Pleysier, 2005; Sireci et al., 2003; Tran & Williams, 1994; van Buuren et al., 2005; Van de Vijver & Poortinga, 2002; Van de Vijver & Leung, 2000; Van Herk et al., 2005; Welkenhuysen-Gybels & Van de Vijver, 2001). Bag ønsker om statistiske ækvivalenstest ligger forbehold over for "... uncritical acceptance of observed differences in the social domain as reflecting valid cross-cultural differences" (Van de Vijver & Leung, op.cit.:35; tilsvarende, se fx Mullen, 1995).

Fra den righoldige litteratur om ækvivalenstest skal enkelte eksempler nævnes. I en artikel udvikler Blasius & Thiessen (2006) en kvantitativ procedure, der gør det muligt at vurdere holdbarheden af forudsætningen om strukturel ækvivalens (se afsnit 2.3.). På grundlag af ISSP-data fra 1994 og omfattende 24 lande konkluderer forfatterne: "... our analyses reveal that both the underlying construct space and the quality of data are highly uneven between countries, limiting their comparability" (op.cit.:230). Data, der ikke kan sammenlignes, anbefales ekskluderet.

En anden artikel sætter fokus på statistiske test af konstruktionsækvivalens til sikring af, at respondenter med forskellige kulturelle tilhørsforhold forbinder det samme med et givet begreb (Welkenhuysen-Gybels og Van de Vijver, 2001). Forskellige statistiske metoder præsenteres, hvilket typisk vil sige parvise sammenligninger af faktorer eller dimensioner på tværs af kulturelle

---

grupper. Forskerne retter navnlig opmærksomheden imod såkaldt eksploratorisk faktoranalyse, der kan gennemføres ved anvendelse af SAS eller SPSS (ang. eksploratorisk faktoranalyse, se fx også Van de Vijver & Poortinga, 2002).

Yderligere en artikel bygger på en Rasch-model, hvis anvendelse bidrager til sammenlignelige analyser på grundlag af ikke-sammenlignelige multisurveydata (van Buuren et al., 2005): "The method, termed Response Conversion (RC), addresses the problem of divergent formulations of survey questions. RC attempts to transform responses obtained on different items (questions) onto a common scale. Where this can be done, comparisons can be made using the common scale. The method consists of two steps. The first step is the construction of a conversion key by means of a statistical model, in our case the polytomous Rasch model (...). The second step uses the conversion key to convert the information onto the common scale. This step is straightforward, and can be repeatedly done on a routine basis as new information arrives".

Et sidste eksempel angår test af måleenheds- og skalaækvivalens i forbindelse med oversatte spørgeskemaer (se afsnit 2.3): "To facilitate measurement equivalence across different language versions of an assessment, great care must be taken in adapting the instruments and statistical analyses must be conducted to evaluate the comparability of final products" (Sireci et al., 2003:130). Mere konkret handler det om følgende metoder: "In this study, we used weighted multidimensional scaling (MDS), analysis of covariance (ANCOVA), and ordinal logistic regression (LR) to evaluate the structural equivalence and differential item functioning (DIF) of an employee attitude survey from a large international corporation" (op.cit.; ang. MDS, se fx også Braun & Scott, 1998).

## SAMMENFATNING

### 6.1. INDLEDNING

De seneste årtier har været kendetegnet ved en vækst i antallet af multisurveys, dvs. surveyundersøgelser med fokus på sammenligning af lande, regioner eller kulturer. I betragtning af globalisering og øget europæisk integration er væksten forventelig.

I sociologi og tilgrænsende samfundsvidenskaber har komparativ fokusering en central placering – og har haft det siden sociologiens oprindelse, fx i Durkheims klassiske undersøgelser af selvmord. Selv om multisurveys i princippet blot er flere simultane surveyundersøgelser gennemført i forskellige lande og/eller kulturer, synes de umiddelbart at adskille sig fra monosurveys på flere måder, fx med hensyn til følgende:

- komparativ fokusering
- datasammenlignelighed
- begreber og begrebsdefinitioner
- respondenters opgaveløsninger.

Multisurveys er i sigens natur kendetegnet ved særlig udtalt *komparativ fokusering*, hvis problemer har været kendt i årtier. Almindeligvis er monosurveys imidlertid også analytisk sammenlignende, fx ved at påpege sociale eller andre ligheder og forskelle mellem forskellige befolkningsgrupper. Det komparative aspekt er således ikke et særegent træk ved multisurveys. Der er snarere tale om gradforskelle med hensyn til den rolle, som komparative analyser spiller i mono- og multisurveys.

Komparativ fokusering i multisurveys angår både tværnational og ”tværkulturel” surveyforskning. Når surveys er tværnationale, retter analyser sig imod ligheder og forskelle lande imellem, fx ved at undersøge sociale fænomener i to eller flere lande med sigte på sammenligning af, hvordan og i hvilket omfang de manifesterer sig i forskellige kontekster. Derimod sætter ”tværkulturelle” surveyundersøgelser selvsagt fokus på kulturspecifikke ligheder og forskelle, men hvad kultur egentlig er, forbliver uklart i multisurvey-sk (metode)litteratur.

Kultur, der hidrører fra det latinske *cultura*, kan fx betyde vaner, menneskelig skaben og resultater heraf, fælles sprog og værdier, livsstil, symbolmiljø, habitus, menings- og relationssystemer. Som fx Schaffer & Riordan (2003) påpeger, er der derfor behov for, at kulturbegrebet afklares og tildeles en tydelig teoretisk placering: ”... the construct of culture should have a theoretical role in the cross-cultural research framework” (p. 176). Denne rapport forfatter gør imidlertid gældende, at der på grund af kulturbegrebets hyperkompleksitet næppe nogensinde vil blive tale om en konsensuel afklaring. Det vil derfor vedvarende være uklart, hvad der egentlig menes, når multisurveys sætter fokus på ”kulturelle” forskelle. I én multisurvey kan kultur fx betyde fælles sociale vaner, mens begrebet i en anden henviser til fælles menings- eller relationssystemer.

Måleproblemer er velkendte i surveyundersøgelser. Gennem talrige undersøgelser og eksperimenter – navnlig amerikanske – er det efterhånden solidt empirisk underbygget, at data fra én og samme survey ikke nødvendigvis honorerer fordringer om *sammenlignelighed* på tværs af forskellige grupper (se fx Olsen, 2005). Desuden øver måleproblemer ikke kun indflydelse på enkeltvariablers gennemsnit og spredning, men kan også generere ikke-formresistente korrelationer, dvs. påvirke sammenhænge mellem variable. I multisurveys synes der imidlertid at opstå yderligere komplikationer, fordi



risikoen for, at forskellige spørgeskemaer og nationalt eller kulturelt forskellige respondenter skaber usammenlignelige data her angiveligt er større: "... the opportunities for error are multiplied many times over when one moves from carrying out research in a single society, to doing it in many different societies" (Inglehart & Welzel, 2004).

På grund af multisurveys komparative fokusering på ligheder og forskelle mellem lande og kulturer følger en særlig opmærksomhed rettet imod datas sammenlignelighed eller mangel på samme. Derfor taler også det voksende antal multisurveys, der selvsagt accentuerer behovet for sammenlignelighed. På den anden side er behovet for fokusering på datasammenlignelighed ikke et særligt kendetegn ved multisurveys (se fx Olsen, 2001a), ligesom det ikke vides, hvor meget større problemer vedrørende sammenlignelighed måtte være i multisurveys. At problemerne er større i multisurveys tages almindeligvis for givet, men er ikke undersøgt.

Yderligere et multisurveysk problem er identifikation af *tværgående begreber* samt definition og operationalisering af disse. Mere eller mindre abstrakte begreber, der indgår i multisurveyske problemstillinger, bør selvsagt være relevante i tværnational og/eller -kulturel henseende, og er desuden ofte anderledes end begreber, som monosurveys tager afsæt i. Selv om også monosurveys som nævnt er komparative, synes multisurveys at rejse særlige konceptuelle komparationsproblemer. Eftersom ikke alle begreber er kulturelt universelle, er det nødvendigt at tage stilling til, om multisurveyske problemstillinger er kulturuniverselle eller kulturspecifikke.

I dele af angelsaksisk litteratur benævnes det universelle perspektiv *etic*, mens det specifikke kaldes *emic*. Mens den "emiske" tilgang så at sige fokuserer "indefra" på specifikke kulturer, har den "etiske" tilgang bredere komparative synsvinkler og retter sig imod forskellige kulturer. Ved gennemførelse af multisurveys bør tilgangen være helt eller i al fald overvejende "etisk", men kan vise sig at være påtvungen, nemlig hvis "emiske" forskelle ikke desto mindre gør sig gældende. *Imposed etics* kan imidlertid også forekomme i monosurveys, fordi begreber heller ikke nødvendigvis er "etiske" der, fx hvis en monosurvey sætter fokus på befolkningsgrupper, der er markant forskellige i kulturel henseende. Eksempelvis er mange SFI-monosurveys baseret på repræsentative nationale stikprøver af voksne, som er bosiddende i Danmark, men som sagtens kan tilhøre forskellige kulturer.

Når stikprøver både omfatter geografiske områder med mange etniske minoriteter (fx Vollsmose) og områder næsten eller helt uden (fx Rungsted), er risikoen for monosurveysk *imposed etics* betydelig.

Endelig er også nationalt og/eller kulturelt forskellige respondenter måde at løse *kognitive svaropgaver* på afgørende for, om mulighederne for statistiske sammenligninger reduceres af måleproblemer eller eventuelt helt falder bort. Selv om det ikke er undersøgt, synes respondenteres svarprocessuelle bidrag til generering af måleproblemer at være særlig udtalt i multisurveys, fx fordi sprog og sprogforståelse, andre kognitive processer og sociale konventioner varierer fra land til land og fra kultur til kultur. Kulturelt betingede forskelle i måder, hvorpå spørgsmål forstås, kan fx føre til data, der tilsyneladende peger på manifesterede kulturelle forskelle, men som skyldes forskelle i respondenteres forståelse af spørgsmål. Fordi noget andet end det forskningsmæssigt tilsigtede måles, opstår der hermed et validitetsproblem.

Hertil kommer, at også forskelle fx med hensyn til respondenteres genkaldelse af informationer, konteksteffekter og/eller grader af redigering af genkaldte informationer kan føre til mangelfulde eller uanvendelige data. Det er derfor en særdeles krævende udfordring at minimere svarprocessuelt betingede måleproblemer i multisurveys (se fx Smith, 2004). Tilsvarende har imidlertid også vist sig at være afgørende i monosurveys. Det gælder fx blandt danske respondenter, hvis forståelse af spørgsmål og genkaldelse af informationer er markant forskellig, fx geografisk (Olsen, 2001a). Også med hensyn til virkninger af løsning af kognitive opgaver forekommer det derfor vanskeligt at hævde, at forskellen mellem multi- og monosurveys skulle være af grundlæggende beskaffenhed.

Ikke desto mindre er nogle multisurveyske forskere af den opfattelse, at den slags surveys forudsætter implementering af en særlig metodologi (se fx Mohler, 2005). Til støtte for den tilsyneladende ret udbredte opfattelse fremføres, at multisurveys er mere komplekse end monosurveys, fx på grund af flere interview og mere omfattende datamateriale, ligesom problemer i forbindelse med kulturelt tværgående begrebsdefinitioner og begrebsoperationalisering fremhæves. Som de indledende overvejelser anskueliggør, kan det dog – ikke kun af rapportens forfatter – gøres gældende, at der næppe er grundlæggende metodologiske forskelle mellem multi- og monosurveys, men udelukkende gradforskelle: ”In terms of issues of reliability, validity, repli-

cability, and generalizability, the comparative study is no different from the cross-sectional design” (Bryman, 2000:52).

## 6.2. PROBLEMATISERENDE SAMMENFATNING

Eftersom et stadigt voksende antal surveyundersøgelser, ofte med dansk deltagelse, forudsætter sammenlignelige data over grænser eller på tværs af kulturer, er der gode grunde til at sætte fokus på de særlige måleproblemer, som sammenligning af data på tværs af surveyundersøgelser rejser, og på procedurer til forebyggelse heraf.

Temaet, der udmøntes i fire problemstillinger, er derfor, hvordan surveydata konstrueres og hvordan datakonstruktion kan kvalitetssikres ved gennemførelse af surveyundersøgelser over grænser og på tværs af kulturer: Hvad er multisurveyske måleproblemer, og hvordan konceptualiseres problemerne? Hvordan formuleres spørgsmål over grænser eller på tværs af kulturer, og hvordan oversættes spørgsmål i multisurveys til forebyggelse af måleproblemer? Hvordan løser respondenter semantiske og andre kognitive opgaver over grænser og på tværs af kulturer? Hvordan afprøves spørgeskemaer til forebyggelse af multisurveyske måleproblemer?

### **Måleproblemers konceptualisering**

Rapportens første problemstilling handler i sidste instans om, hvorvidt multi- og monosurveys er grundlæggende forskellige. Bortset fra at multisurveys overskrider grænser og/eller kulturer og derfor stiller særlige praktiske og andre krav til gennemførelse, fx oversættelse af spørgeskemaer, er der indledningsvis argumenteret for, at der næppe er basale forskelle mellem multi- og monosurveys. I princippet er multisurveys blot flere samtidige surveys gennemført i forskellige lande og/eller kulturer. Ganske vist er komparativ fokusering, sammenlignelighed, begreber og deres definitioner samt respondents vekslende kognitive opgaveløsninger muligvis særlig vigtige multisurveyske aspekter, men også monosurveys er komparative, fx ved at modstille grupperes levkår eller holdninger, ligesom de kræver datasammenlignelighed, begrebsdefinitioner og begrebsoperationaliseringer samt tilfredsstillende udførelse af respondents svararbejde.

Hvis forskellene mellem multi- og monosurveys således kun er gradforskelle, er den afvigende multi- og monosurveyske konceptualisering af kvalitetssikring overraskende (se kapitel 2). I monosurveysk litteratur er der næsten fuldstændig konsensus om validitet og reliabilitet som konceptuelt grundlag for forebyggelse af måleproblemer. På den anden side er der dog ikke enighed om, hvordan begreberne defineres og operationaliseres, ligesom der heller ikke er overensstemmelse med hensyn til subbegreber knyttet til validitetsbegrebet, fx indholdsvaliditet, begrebsvaliditet etc. Endelig kan det være vanskeligt at opretholde sondringen i praksis, hvorfor nogle forskere enhedsbetegner de to begreber under overskriften måleproblem eller *measurement error*.

Multisurveysk metodelitteratur er derimod kendetegnet ved ækvivalensbegrebets dominans. Selv om begrebet ofte forekommer i almen form, defineres det sjældent eller med diffuse formuleringer, der langt fra honorerer følgende krav: "A key foundation is a systematic conceptual framework for survey quality" (Lynn, 2003:335). Derimod defineres en lang række specifikke ækvivalensformer, hvoraf ofte nævnte fx er funktionel, konceptuel og semantisk. Også definitioner af specifikke ækvivalensformer er ofte vage og modsætningsfyldte: "... terms used in the literature to describe equivalence are often unclear and confusing" (Van de Vijver & Leung, i: Berry et al., 1997:261). Desuden synes flere begreber redundante, fx semantisk ækvivalens og oversættelsesækvivalens. Endelig grupperes specifikke ækvivalensbegreber undertiden i overordnede ækvivalens kategorier, fx fortolkende i forhold til procedural ækvivalens, hvor førstnævnte fx omfatter funktionel og konceptuel ækvivalens, mens sidstnævnte fx angår metrisk ækvivalens og måleenhedsækvivalens.

I de fleste fremstillinger virker ækvivalensbegrebet og dets vekslende former som eneste kvalitetssikrende begreb, mens det i andre sammenhænge nævnes i kombination med begreberne validitet og reliabilitet, fx af Johnson (1998): "In addition to the traditional reliability and validity requirements for monocultural survey instruments, researchers conducting cross-cultural survey research have the added concern of equivalence" (p. 30). Ækvivalens virker i så fald som supplerende kvalitetssikrende begreb, hvortil kommer enkelte forfatteres anvendelse af bias-begrebet.

Da multisurveys overskrider grænser eller kulturer, er fokusering på sammenlignelighed nærliggende, men forklarer på den anden side ikke, hvorfor der skulle være behov for et kvalitetsbegreb, der overskrider velkendte og i samfundsvidenskabelige discipliner almindeligvis anvendte begreber som reliabilitet og validitet. Uanset om ækvivalensbegrebet er alenestående eller kombineres med gængse krav om målinger, der honorerer krav om validitet og reliabilitet, burde målinger, der opfylder begge krav, også være sammenlignelige over grænser og kulturer. At målinger kan være pålidelige uden at være gyldige er velkendt, men hvordan målinger kan være valide uden at være ækvivalente, forekommer uforståeligt.

Når ækvivalensbegrebet ikke desto mindre virker som afgørende konceptuelt grundlag for kvalitetssikring af multisurveydata, har det efter alt at dømme hverken videnskabsteoretisk, teoretisk eller metodologisk begrundelse. Intet tyder fx på, at multisurveyforskere indtager anderledes ontologiske og epistemologiske positioner end monosurveyske forskere. At multisurveyforskere anvender andre kvalitetssikrende begreber end deres monosurveykolleger er derfor ikke nødvendigvis udtryk for, at parterne tænker særlig forskelligt. Begrundelsen er snarere af pragmatisk beskaffenhed. Ved at tildele ækvivalensbegrebet forrang ønsker multisurveyforskere at gøre hinanden og det videnskabelige samfund opmærksom på multisurveys udtalte komparative fokusering og de særlige forskningsmæssige og praktiske udfordringer, der følger heraf.

### **Spørgeskemakonstruktion og oversættelse**

Rapportens anden problemstilling er, hvordan anvendelige spørgsmål formuleres over grænser og på tværs af kulturer, og dermed også hvordan spørgsmål oversættes. Mens monosurveysk metodelitteratur er righoldig, når det handler om forslag til, hvordan spørgsmål formuleres, svarkategorier udarbejdes og spørgeskemahelheder dannes, gælder det ikke multisurveysk. I så henseende er langt de fleste anbefalinger overtaget direkte fra monosurveysk litteratur, fx vedrørende spørgsmåls fokusering, længde, formulering og anvendelse af præciserende *probes*.

For så vidt er anbefalingerne vedrørende forskellige former for national og/eller kulturel tilpasning af spørgsmål heller ikke specifikt multisurveyske, eftersom samme problematik gør sig gældende i monosurveys. Også

i sidstnævnte er det nødvendigt at tilpasse spørgsmål til foreliggende sociale, retslige og andre konkrete betingelser. Kun anbefalingen af oversættelseseg-nethed er ny. Hvis multisurveys virkelig er så metodologisk forskellige fra monosurveys, som mange forskere gør gældende, er det overraskende, at der stort set ikke er gjort forsøg på – med afsæt i den omfangsrige monosur-veyske metodelitteratur – at reflektere over og udvikle ledetråde for, hvilke eventuelle særlige problemer multisurveysk spørgeskemakonstruktion måtte rejse.

Bortset fra overvejelserne om kulturuniverselle ”etiske” (*etic*) versus kulturspecifikke ”emiske” (*emic*) begreber og indikatorer er der for så vidt heller ikke, når det handler om operationaliserende at bevæge sig ned ad ”abstraktionsstigen” fra abstrakte begreber til empiriske begrebsindikatorer, tale om anbefalinger, der afviger fra monosurveysk litteratur. Også mono-surveys forudsætter nemlig konceptuel sammenlignelighed, ligesom anbefa-lingen af at anvende to eller flere empiriske indikatorer på samme begreb er velkendt fra monosurveysk metodelitteratur. Det gælder ikke mindst, når befolkningers holdninger og værdier måles.

Når landegrænser eller kulturer overskrides, øges risikoen efter alt at dømme for, at abstrakte begreber, som surveys baseres på, ikke fører til ækvivalente målinger. Derfor er litteraturen om ”etiske” versus ”emiske” begreber og om *imposed etics* særdeles relevant, hvilket dog ikke kun gælder i multisurveyske sammenhænge, men også i monosurveyske, hvor sproglige og kulturelle forskelle undertiden er langt større, end surveyforskere umiddel-bart formoder. Det gælder bl.a. i en *split sample*-baseret undersøgelse blandt voksne danskere, der fx er markant forskellige med hensyn til hvordan ”samme” spørgsmål forstås (Olsen, 2001a).

Litteraturens helt særlige – fra monosurveyske fremstillinger afvi-gende – bidrag handler imidlertid om oversættelse af spørgeskemaer. Selv om litteraturen, der kun baseres på enkelte empiriske undersøgelser, langt fra er af konsensuel beskaffenhed, indeholder den adskillige refleksioner og anbefalinger, der ikke blot er nyttige ved gennemførelse af multisurveys, men også i tilfælde, hvor monosurveyske spørgeskemaer påtænkes oversat, fx når indvandrere og flygtninge skal interviewes. Det gælder generelle ledetråde, fx angående semantisk ækvivalens, men også en række mere specifikke oversæt-telsesprocedurale anbefalinger, fx vedrørende kvalifikationer og samarbejde.

Der er formodentlig tale om anbefalinger, som er ukendte for ganske mange danske surveyforskere.

### **Respondenters opgaveløsninger**

Rapportens tredje problemstilling er, hvordan respondenter løser kognitive opgaver over grænser og på tværs af kulturer. Det drejer sig om respondenters vekslende forståelse af spørgsmål og svarkategorier, genkaldelse af ønskede informationer, konteksteffekter og ”socialt ønsket” redigering af genkaldte informationer, før svar afgives.

Når det handler om, hvordan forståelse af spørgsmål og svarkategorier skaber måle- eller ækvivalensproblemer, er monosurveysk metodeforskning igen omfattende, mens multisurveysk er mere sporadisk. Mens monosurveysk forskning omhandler en lang række problematikker fra forståelse af ord over spørgsmålssemantik til adskillige svarkategorielle aspekter, er multisurveysk metodeforskning mindre omfangsrig og tematisk snævrere. Det gælder også forskningsbaserede anbefalinger med henblik på reduktion af ækvivalensproblemer. Adskillige undersøgelser peger dog på, at variationer i ”grænseoverskridende” forståelse af ord, spørgsmål og svarkategorier skaber ækvivalensproblemer. Intet tyder imidlertid på, at der skulle foreligge en grundlæggende anderledes metodeproblematik, men muligvis større måleproblemer. Johnson et al. (1997) og andre mener, at kulturelle forskelle skaber yderligere sammenligningsproblemer, men forskningen er for mangelfuld til at kunne vurdere dette.

Med hensyn til respondenters genkaldelse af informationer er monosurveysk metodeforskning ligeledes omfattende og bredspektret, mens multisurveysk er endnu mere sporadisk end vedrørende spørgsmålsforståelse. Mens monosurveysk forskning fx både omhandler, hvordan genkaldelse, glemsel, teleskopering, visse indkodede informationers ikke-karakteristiske træk og kognitiv magelighed genererer måleproblemer, er det multisurveyske fokus smallere. Kun angående strategier for genkaldelse af informationer samt kognitiv magelighed findes få ”grænseoverskridende” resultater. Hvad kognitiv magelighed angår, drejer det sig især om et beslægtet aspekt, nemlig føjelighed. Alt i alt forbliver det derfor uafklaret, hvornår og under hvilke betingelser respondenters genkaldelse af informationer er kulturelt sensitiv. Som bl.a. påpeget af Johnson et al. (1997) er der derfor et betydeligt forsk-

ningsbehov, der fx kan tilgodeses ved at overføre CASM-erfaringer til multi-surveyesk metodeforskning.

Når spørgsmålsforståelse eller andre svarprocessuelle opgaver afhænger af tidligere spørgsmål, kan målinger være behæftet med kontekstuel betingede fejl. Herom findes en omfattende monosurveyesk litteratur, der fx handler om forskellige typer af konteksteffekter og deres udbredelse, men som dog ikke udpeger blot nogenlunde sikre procedurer til forebyggelse af den slags effekter. På den anden side og værre er tilsvarende multisurveyesk forskning stort set fraværende. Når konteksteffekter har vist sig at generere måleproblemer i monosurveys, bør antagelsen ifølge Smith (i: Harkness et al., 2003) og andre – det gælder også rapportens forfatter – være, at effekterne i al fald ikke er af mindre omfang, når grænser eller kulturer overskrides. Der er derfor også et betydeligt forskningsbehov med hensyn til forebyggelse af kontekstuelle ækvivalensproblemer: "... coordinated, cross-national, experimental studies are lacking" (op.cit.:84).

Hvad endelig svarredigering angår, peger monosurveyeske undersøgelser på, at mange respondenter ofte redigerer genkaldte informationer, før svar afgives, eksempelvis fordi formidlede oplysninger, der er "socialt uønskede", opleves som potentielt stigmatiserende. Hvis den slags dermed genererede måleproblemer skal forebygges, er sikring af anonymitet vigtig. Dette påpeges ligeledes i multisurveyesk forskning, der er ret omfattende, når det handler om undersøgelser af, hvordan kulturelt vekslende *social desirability* blandt respondenter kan generere ækvivalensproblemer. I de fleste undersøgelser peges på kulturelt betingede svarredigerende variationer, hvorfor redigering er et problem, der bør tages højde for ved gennemførelse af multisurveys, fx ved at formulere så lidt "truende" spørgsmål som muligt. Det ligger derfor lige for, at forskere, der gennemfører multisurveys, bør være i besiddelse af og kunne dokumentere tværkulturel forståelse og indsigt.

### **Afprøvning af spørgeskemaer**

Rapportens fjerde problemstilling er, hvordan spørgeskemaer afprøves til forebyggelse af måleproblemer i multisurveys. I forlængelse af belysningen af rapportens første problemstilling angår én væsentlig forskel kvalitets-sikrende spørgeskemaafprøvnings konceptuelle grundlag, hvor monosurveys ifølge metodelitteraturen skal honorere krav om validitet og reliabilitet,



mens den gennemgående fordring som nævnt er ækvivalens eller sammenlignelighed i multisurveysk litteratur. Hvorfor afprøvning af validitet og reliabilitet ikke er tilstrækkelig, besvarer multisurveysk litteratur ikke. Det forbliver med andre ord uklart, hvad ækvivalenssikring bidrager med, hvis surveydatas validitet og reliabilitet allerede er maksimalt sikret.

Yderligere en væsentlig forskel er udtalt mangel på empiriske undersøgelser af, hvordan prætest virker, og hvor effektive de måtte være i multisurveyske sammenhænge. Indtil sådanne undersøgelser foreligger, foreslår toneangivende forskere tillige med rapportens forfatter en række almene ledetråde, der sammenfattes af Smith (i: Presser et al. 2004), hvad tværnationale surveys angår, men som kan overføres til tværkulturelle:

1. the multipretesting procedures should be carried out across countries and languages with the results evaluated by researchers expert in
  - a. the cultures and languages being investigated,
  - b. the substantive domains being studied, and
  - c. survey research methodology;
2. the pretesting and translation should be integrated and interactive processes;
3. the pretesting needs to be cross-national and not just within specific countries; and
4. the developmental process takes much more time and resources than for single-country monolingual studies and usually should involve multiple rounds of pretesting and larger samples (p. 451f).

Ofte i monosurveysk litteratur nævnte procedurer til afprøvning af spørgeskemaer er prøveundersøgelser, fokusgrupper, kognitive laboratorieinterview, interaktionskodning og *split sample*-eksperimenter. Også multisurveysk metodelitteratur sætter fokus på prøveundersøgelser, der efter enkeltes opfattelse kan reducere ækvivalensproblemer, men almindeligvis menes andre procedurer mere anbefalelsesværdige, fx kognitive laboratorieinterview. Ligesom en del monosurveyske forfattere påpeger også multisurveyske konventionelle prøveundersøgelses ulemper, men på den anden side er den slags prætest bedre end intet at gøre. Hvad fokusgrupper angår, anbefales de i begrænset omfang multisurveysk. Det kan bl.a. skyldes, at fokusgruppedrøf-

telser ofte finder sted til præcisering af problemstillinger og i datatilvirkningens indledende fase, mens tværnationale surveys ofte gennemføres ved at ”overføre” et masterspørgeskema fra ét land og ét sprog til andre. Men da selv ”gode” oversættelser ikke nødvendigvis ”fungerer” i praksis, er afprøvning heraf afgørende og kan fx finde sted ved gennemførelse af fokusgruppediskussioner, mener flere metodeforfattere.

Derimod anbefales kognitive laboratorieinterview ofte i multisurveysk litteratur. Ifølge flere forskere – også rapportens forfatter – er den slags tænke-højt-interview velegnede til belysning af, hvordan nationalt eller kulturelt forskellige respondenter løser kognitive svaropgaver. Det gælder også afprøvning af oversatte spørgeskemaer. Det beklages dog, at kognitive interview og analyser heraf endnu er en multisurveysk sjældenhed, men det er samtidig en udbredt forventning, at proceduren efterhånden vil finde voksende multisurveysk udbredelse. Som i monosurveysk litteratur er der dog også kritiske røster, fx om planlægning og gennemførelse af kognitive interview. Relevante informanter skal fx selekteres, resultater kan ikke generaliseres, resultater kan være mangelfulde, og der kan være problemer med at finde kvalificerede interviewere.

I multisurveysk litteratur nævner en del forfattere interaktionskodning som en mulig procedure til reduktion af ækvivalensproblemer, men kun få anbefaler fremgangsmåden, nok bl.a. fordi den indskrives sig i en stimulus-respons-opfattelse, der er vanskelig at praktisere, når grænser eller kulturer overskrides. Det problematiseres fx, om samme adfærds-koder er forventelige på tværs af kulturer, og om respondenter, der deltager i multisurveys, svarer på måder, der er sammenlignelige. Endnu færre nævner *split sample*-eksperimenter. At proceduren omtales af få og anbefales af endnu færre skyldes næppe, at den principielt skulle være ubrugelig i multisurveys, men snarere en række praktiske og tidsmæssige problemer i tilknytning til *split sample*-eksperimenter, der overskrider grænser eller kulturer.

Ud over de nævnte kvalitetssikrende procedurer er også debriefing af respondenter, monitorering af interviewere, debriefing af interviewere og ekspertvurderinger potentielt anvendelige ifølge monosurveysk litteratur, blandt hvilke multisurveyske forfattere navnlig advokerer for brug af ekspertvurderinger af spørgeskemaer og oversættelser. Ifølge litteraturen er det dog en betingelse, at eksperter har specialviden om de sprog og kulturer,

som en given multisurvey sætter fokus på. Mens nogle forfattere anbefaler brug af eksperter ved afprøvning af spørgeskemaer, opfordrer andre til at konsultere eller direkte samarbejde med eksperter i hele det samlede undersøgelsesforløb, fx for at forebygge etnocentriske tendenser.

### **Tværfagligt forskningsbehov**

Af den problematiserende sammenfatning følger et markant behov for multisurvey-sk metodeudvikling med hensyn til forskellige aspekter af konstruktion og afprøvning af ”grænseoverskridende” spørgeskemaer, fx:

- Komparative analyser af kvalitetssikrings konceptuelle grundlag med særlig henblik på ækvivalensbegrebets relevans.
- Analyser af multisurvey-sk spørgeskemakonstruktionens eventuelle specifikke problemer ud over oversættelse af spørgeskemaer.
- Empiriske undersøgelser af fordele og ulemper ved forskellige oversættelsesstrategier.
- Empiriske undersøgelser af, om respondenters opgaveløsninger skaber større måleproblemer i multi- end i monosurvey-sk.
- Empiriske undersøgelser af, hvor kulturelt sensitiv forskellige respondenters genkaldelse af informationer er.
- Empiriske undersøgelser af multisurvey-ske konteksteffekter.
- Empiriske undersøgelser af fordele og ulemper ved forskellige måder, hvorpå multisurvey-ske spørgeskemaer kan afprøves.

Som påpeget gentagne gange i multisurvey-sk metodelitteratur – fx af Harkness (se fx 1999; 2005) – er der derudover et behov for et øget tværfagligt samarbejde, fx inspireret af CASM-bevægelsen (*Cognitive Aspects of Survey Methodology*) og andre, der i de senere år har ydet værdifulde bidrag til monosurvey-sk metodeudvikling. Det drejer sig ikke mindst om samarbejde med lingvister, fx socio-semantikere: ”Linguistics can be expected to offer an ongoing refinement of our understanding of different types of meaning” (Harkness et al., i: Presser et al., 2004a:472). Men også kognitionspsykologer, kommunikationsforskere, antropologer m.fl. kan bidrage til yderligere metodeudvikling, mener flere multisurvey-ske forfattere såvel som den forfatter, der hermed afslutter rapporten.

## LITTERATUR

Litteratur markeret med stjerne (\*) er fra multisurveysk litteratursøgning efteråret 2006, mens litteratur uden stjerne er fra Olsen (2005).

- Aiken, L.R. (1997): Questionnaires and inventories: Surveying opinions and assessing personality. New York: John Wiley.
- Alba, J.W. & Hasher, L. (1983): Is Memory Schematic? *Psychological Bulletin*. Vol. 93, No. 2.
- Aldridge, A. & Levine, K. (2001): Surveying the Social World: Principles and Practice in Survey Research. Philadelphia: Open University Press.
- Allén, S. et al. (1989)\*: Orden speglar samhället. Stockholm: Almänna Förlaget.
- Alreck, P.L. & Settle, R.B. (1995): The Survey Research Handbook. Chicago: Irwin Publications.
- Alwin, D.F. (1977): Making Errors in Surveys. *Sociological Methods & Research*. Vol. 6, No. 2.
- Alwin, D.F. (1991): Research on Survey Quality. *Sociological Methods & Research*, vol. 20, no. 2.

- Alwin, D.F. et al. (1994)\*: Measurement in Multi-National Surveys. I: Borg, I. & Mohler, P.P., Trends and Perspectives in Empirical Research. Berlin: Walter de Gruyter.
- Alwin, D.F. et al. (2002)\*: Problems of Functional Equivalence of Measurements in Multinational Surveys. I: de Vaus, D. (ed.), *Social Surveys*. Vol. III: 293-302. London: Sage Publications.
- Andersen, B.H. & Christoffersen, M.N. (1982): Om spørgeskemaer. København: Socialforskningsinstituttet.
- Andersen, R. et al. (1979): Total Survey Error. San Francisco: Jossey-Bass.
- Andersson, B-E. (1985): Som man frågor får man svar. Stockholm: Liber.
- Andrews, F. (1984): Construct Validity and Error Components of Survey Measures. *Public Opinion Quarterly*. Vol. 48, No. 2.
- Arksey, H. & Knight, P. (1999): Interviewing for social scientists. Thousand Oaks: Sage.
- Babbie, E.R. (1990): Survey Research Methods. Belmont: Wadsworth Publications.
- Baddeley, A. (1990): Human Memory. Theory and Practice. Boston & London.
- Baker, C. (1992): Attitudes and Language. Clevedon: Multilingual Matters Ltd.
- Baldursson, E.B. & Pedersen, B.T. (1992): Fortolkning og mening. *Nordisk Psykologi*. 1992/1.
- Bateson, N. (1984): Data Construction in Social Surveys. London.
- Baumgartner, H. & Steenkamp, J.B. (2001)\*: Response Styles in Marketing Research. *Journal of Marketing Research*, 38: 143-156.
- Beatty, P. (2004)\*: The Dynamics of Cognitive Interviewing. I: Presser, S. et al. (eds.), Methods for Testing and Evaluating Survey Questionnaires. New York: Wiley.
- Behling, P. & Law, K.S. (2000)\*: Translating Questionnaires and Other Research Instruments. California: Sage Publications.
- Belson, W.A. (1968): Respondent Understanding of Survey Questions. *Polls*. Vol. 3.
- Belson, W.A. (1981): The Design and Understanding of Survey Questions. London: Gower.
- Belson, W.A. (1986): Validity in Survey Research. New York: Gower.

- Bergman, L.R. & Wärneryd, B. (1982): Om datainsamling i surveyundersökningar. Stockholm: Statistisk Centralbyrå.
- Bernard, H.P. et al. (1984): On the validity of retrospective data: The problem of informant accuracy. *Annual Review of Anthropology*, 1984/13:495ff.
- Berry, J.W. (1969)\*: On Cross-Cultural Comparability. *International Journal of Psychology*, Vol. 4, pp. 199-228.
- Berry, J.W. (1989)\*: Imposed Etics-Emics-Derived Etics: The Operationalization of a Compelling Idea. *International Journal of Psychology*, Vol. 24:721-735.
- Berry, J.W. (1999)\*: Emics and Etics: A Symbiotic Conception. *Culture and Psychology*, 5: 165-171.
- Berry, J.W. et al. (1997)\*: Handbook of Cross-Cultural Psychology. *Theory and Method*. Vol. 1. Boston: Allyn and Bacon.
- Bickman, L. & Rog, D.J. (eds.) (1998): Handbook of Applied Social Research Methods. Thousand Oaks: Sage.
- Biemer, P.P. et al. (eds.) (1991)\*: Measurement Errors in Surveys. New York: John Wiley & Sons.
- Billiet, J. et al. (2002)\*: Equivalence of Measurement Instruments for Attitude Variables in Comparative Surveys, Taking Method Effects into Account (<http://mrvar.fdv.unilj.si/pub/mz/mz18/billet.pdf>).
- Billiet J. & Welkenhuysen-Gybels, J. (2004)\*: Assessing cross-national construct equivalence in the ESS. Sixth Int. Conference on Social Science Methodology. Amsterdam ([http://naticent02.uuhost.uk.uu.net/publicity/construct\\_equivalence\\_religious\\_involvement.pdf](http://naticent02.uuhost.uk.uu.net/publicity/construct_equivalence_religious_involvement.pdf))
- Billiet, J. & Welkenhuysen-Gybels, J. (2004a)\*: Assessing cross-national construct equivalence in the ESS: the case of six immigration items. I: van Dijkum, C. et al. (Eds.), *Recent Developments and Applications in Social Research Methodology*. Amsterdam: Barbara Budrich Publ. ([http://naticent02.uuhost.uk.uu.net/publicity/construct\\_equivalence\\_immigration.pdf](http://naticent02.uuhost.uk.uu.net/publicity/construct_equivalence_immigration.pdf)).
- Birbili, M. (2000)\*: Translating from one Language to Another. Social Research Update, 31. (<http://www.soc.surrey.ac.uk/sru/SRU31.html>)

- Blair, E. & Burton, S. (1987): Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions. *Journal of Consumer Research*. Vol. 14.
- Blair, J. & Presser, S. (1993)\*: Survey Procedures for Conducting Cognitive Interviews to Pretest Questionnaires. Proceedings of the Section on Survey Research Methods of the American Statistical Association ([http://www.amstat.org/sections/srms/Proceedings/papers/1993\\_059.pdf](http://www.amstat.org/sections/srms/Proceedings/papers/1993_059.pdf)).
- Blais, A. & Gidengil, E. (1993)\*: Things are not always what they seem: French-English Differences and the Problem of Measurement Equivalence. *Canadian Journal of Political Science*. Vol. 26: 541-555.
- Blasius, J. & Thiessen, V. (2006)\*: Assessing Data Quality and Construct Comparability in Cross-National Surveys. *European Sociological Review*. Vol. 22, No. 3: 229-242.
- Blomqvist, J. & Bruce, G. (red.) (1997)\*: Språket och tiden. Lund: Studentlitteratur.
- Borg, I. & Mohler, P.Ph. (1994): Trends and Perspectives in Empirical Social Research. Berlin & New York: Walter de Gruyter.
- Borrie, W.T. & Roggenbuck, J.W. (1998): The problem of verbal reports in recreation research: Review, recommendation, and new directions. University of Montana, School of Forestry, Missoula ([www.metabase.net/docs/iht/01827.html](http://www.metabase.net/docs/iht/01827.html)).
- Bourque, L.B. & Clark, V.A. (1992): Processing Data: The Survey Example. Newbury Park: Sage Publications.
- Bourque, L.B. & Fielder, E.P. (1995): How to Conduct Self-Administered and Mail Surveys. Thousand Oaks: Sage Publications.
- Bower, G.H. et al. (1979): Scripts in Memory for Text. *Cognitive Psychology*. Vol. 11.
- Bradburn, N.M. (2000): Questionnaire Design: from Art into Science. Fifth Int. Conference on Social Science Methodology Cologne, Germany. National Science Foundation (<http://cloud9.norc.uchicago.edu/-dlib/colog.htm>).
- Bradburn, N.M. et al. (1987): Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys. *Science*. Vol. 236, No. 10.

- Bradburn, N.M. & Mason, W.M. (1964): The Effect of Question Order on Responses. *Journal of Marketing Research*. November.
- Bradburn, N.M. & Miles, C. (1979): Vague Quantifiers. *Public Opinion Quarterly*. Vol. 43, No. 1.
- Bradburn, N.M. & Sudman, S. (1979): Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass Publ.
- Bradburn, N.M. & Sudman, S. (1989): Polls & Surveys: Understanding what they tell us. San Francisco: Jossey-Bass.
- Brannen, J. (ed.) (1992): Mixing Methods: Qualitative and Quantitative Research. Avebury: Aldershot.
- Braun, M. & Scott, J. (1996)\*: Data-based procedures for the deception of problems in functional equivalence. Int. Sociological Association Conference on Social Science Methodology. University of Essex, England.
- Braun, M. & Scott, J. (1998)\*: Multidimensional Scaling and Equivalence: *Is having a job the same as working?* ZUMA-Nachrichten Spezial, January 1998: 129-144. ([http://www.gesis.org/publikationen/Zeitschriften/ZUMA\\_Nachrichten\\_spezial/documents/znspezial3/znspez3\\_05\\_Braun\\_Scott.pdf](http://www.gesis.org/publikationen/Zeitschriften/ZUMA_Nachrichten_spezial/documents/znspezial3/znspez3_05_Braun_Scott.pdf)).
- Braverman, M.T. & Slater, J.K. (eds.) (1996): Advances in Survey Research. San Francisco: Jossey-Bass.
- Brenner, M. et al. (eds.) (1985): The Research Interview. Uses and Approaches. London & New York: Academic Press.
- Briggs, C.L. (1986): Learning How to Ask: A Sociolinguistic Appraisal of the Role of the Interview in Social Science Research. Cambridge University Press.
- Brislin, R.W. (1970)\*: Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*. No. 1: 185-216.
- Brown, N.R. et al. (1985): The Subjective Dates of Natural Events in Very-Long-Term Memory. *Cognitive Psychology*. Vol. 17.
- Bryman, A. (2000)\*: Social Research Methods. Oxford: University Press.
- Bullinger, M. (2003)\*: International comparability of health interview surveys. I: Nosikov, A. & Gudex, C. (eds.), Developing Common Instruments for Health Surveys (<http://www.euro.who.int/document/HIS/004Chapt01.pdf>).



- Bulmer, M. (1998)\*: The problem of exporting social survey research. *American Behavioral Scientist*. Vol. 42: 153-167.
- Burns, R.B. (1994): Introduction to Research Methods. Melbourne: Longman Cheshire.
- Burton, S. & Blair, E. (1991): Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions. *Public Opinion Quarterly*. Vol. 55.
- Byrne, D. (2002): Interpreting Quantitative Data. London: Sage Publications.
- Cantril, H. (1944): Gauging Public Opinion. London: Princeton.
- Carrasco, L. (2003)\*: The American Community Survey en Español. Washington D.C.: Statistical Research Division.
- Chen, C. et al. (1995)\*: Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6: 170-175.
- Cheng, A.T.A. (2001)\*: Case definition and culture: are people all the same? *The British Journal of Psychiatry*, 179: 1-3.
- Cheung, G.W. & Rensvold, R.B. (2000)\*: Assessing Extreme and Acquiescence Response Sets in Cross-Cultural Research Using Structural Equations Modeling. *Journal of Cross-Cultural Psychology*. Vol. 31, No. 2: 187-212.
- Cicourel, A.V. (1964): Method and Measurement in Sociology. London.
- Cicourel, A.V. (1973): Cognitive Sociology. *Language and Meaning in Social Interaction*. Middlesex.
- Cicourel, A.V. (1981): The Role of Cognitive-Linguistic Concepts in Understanding. *Annual Review of Sociology*. Vol. 7.
- Clark, H.H. & Schober, M.F. (1992)\*: Asking Questions and Influencing Answers. I Tanur, J. (ed.): Questions about Questions. New York: Springer.
- Clarke, I. (2001)\*: Extreme response style in cross-cultural research. *International Marketing Review*. Vol. 18, No. 3: 301-324.
- Cohen, S.B. & Lepkowski, J.M. (2004)\*: Health Survey Research Methods. Maryland (<http://www.cdc.gov/nchs/data/misc/proceedings.hsr-2004.pdf>).
- Converse, J.M. & Presser, S. (1986): Survey Questions: Handcrafting the Standardized Questionnaire. Beverly Hills: Sage.

- Cool, C., & Spink, A. (2002)\*: Issues of Context in Information Retrieval: An Introduction to the Special Issue. *Information Processing & Management*. Vol. 38, No. 5: 605-11.
- Couper, M.P. et al. (eds.) (1998): Computer Assisted Survey Information Collection. Wiley.
- Crano, W.D. & Brewer, M.B. (2002): Principles and Methods of Social Research. Lawrence Erlbaum Ass.
- Crotty, M. (1998): The Foundations of Social Research. London: Sage.
- Czaja, R. (1998): Questionnaire Pretesting Comes of Age. *Marketing Bulletin*, 1998/9:52-66.
- Czaja, R. & Blair, J. (1996): Designing surveys: A guide to decisions and procedures. Thousand Oaks: Pine Forge Press.
- Dahmström, K. (1996): Från datainsamling till rapport: Att göra en statistisk undersökning. Lund: Studentlitteratur.
- Davis, D.W. (1997)\*: Nonrandom Measurement Error and Race of Interviewer Effects. *Public Opinion Quarterly*. Vol. 61: 183-207.
- DeMaio, T.J. et al. (2003): Improving survey quality through pretesting. U.S. Bureau of the Census ([www.census.gov/srd/papers/pdf/sm98-03.pdf](http://www.census.gov/srd/papers/pdf/sm98-03.pdf)).
- de Vaus, D.A. (1998): Surveys in Social Research. London: UCL Press.
- de Vaus, D.A. (2002): Surveys in Social Research. London: UCL Press.
- de Vaus, D.A. (ed.) (2002a): Social surveys (bd. I-IV). London: Sage.
- Dijkstra, N. & van der Zouwen, J. (1982): Response Behaviour in the Survey Interview. London: Academic Press.
- Douglas, S.P. & Nijssen, E.J. (2002)\*: On the Use of 'Borrowed' Scales in Cross-National Research. University of Nijmegen (<http://pages-stern.nyu.edu/~sdouglas/rpubs/borrowed.doc>).
- Drew, J.H. (1991): Modelling the Effects of Question Order and Form. American Statistical Association. Proceedings of the Section on Survey Research Methods. Atlanta, Georgia.
- Eagly, A.H. & Chaiken, S. (1993): The Psychology of Attitudes. New York: Harcourt Brace College Publ.
- Edenborough, R. (2002): Effective interviewing. London: Kogan Page.
- Eiser, J.R. (1986): Social Psychology. Attitudes, Cognition and Social Behavior. Cambridge University Press.

- Ember, C.R., & Ember, M. (2001)\*: Cross-cultural research methods. Walnut Creek: AltaMira.
- Eneroth, B. (1984): Hur mäter man "vackert"? Stockholm: Liber.
- Ercikan, K. (1998)\*: Translation Effects in International Assessments. *International Journal of Educational Research*. Vol. 29: 543-553.
- Esposito, J.L. et al. (1991): Determining Which Questions Are the Best. American Statistical Association. Proceedings of the Section on Survey Research Methods. Atlanta, Georgia.
- Eysenck, M.W. (1993): Principles of Cognitive Psychology. Hove (UK) & Hillsdale (USA).
- Eysenck, M.W. & Keane, M.T. (1995): Cognitive Psychology. Hove (UK) & Hillsdale (USA).
- Fee, J.F. (1981): Symbols in Survey Questions: Solving the Problem of Multiple Word Meaning. *Political Methodology*. No. 7.
- Fienberg, S.E. & Tanur, J.M. (1989): Combining Cognitive and Statistical Approaches to Survey Design. *Science*. Vol. 243.
- Fink, A. (1995a): How to Sample in Surveys. Thousand Oaks: Sage Publications.
- Fink, A. (1995b): How to Report on Surveys. Thousand Oaks: Sage Publications.
- Fink, A. (1995c): How to Design Surveys. Thousand Oaks: Sage Publications.
- Fink, A. (1995d): How to Ask Survey Questions. Thousand Oaks: Sage Publications.
- Fink, A. (1995e): How to Analyze Survey Data. Thousand Oaks: Sage Publications.
- Fink, A. (1995): The Survey Handbook. Thousand Oaks: Sage Publications.
- Fink, A. & Kosecoff, J. (1998): How to Conduct Surveys: A Step-by-Step Guide. Thousand Oaks: Sage Publications.
- Fink, A.S. (2003): Andet end tallenes tale - Om at se spørgeskemaundersøgelsen som diskurs. Odense: Dansk Data Arkiv ([www.sociologi.dk/db/03papers/58.doc](http://www.sociologi.dk/db/03papers/58.doc)).
- Fischer, R. (2004)\*: Standardization to Account for Cross-Cultural Response Bias. *Journal of Cross-Cultural Psychology*. Vol. 35, No. 3: 263-282.

- Flaherty, J.A. et al. (1988)\*: Developing instruments for cross-cultural psychiatric research. *Journal of Nervous and Mental Disease*, 167: 257-263.
- Flavell, J.H. (1977): *Cognitive Development*. New Jersey.
- Foddy, W. (1993): *Constructing Questions for Interviewing and Questionnaires*. Cambridge: Cambridge University Press.
- Folz, D.H. (1996): *Survey Research for Public Administration*. Thousand Oaks: Sage Publications.
- Foreman, E.K. (1991): *Survey Sampling Principles*. New York: Dekker.
- Forsyth, B.H. (2006)\*: Improving Questionnaire Translations and Translation Processes. Proceedings of Q2006. European Conference on Quality in Survey Statistics. ([http://www.statistics.gov.uk/events/-q2006/downloads/T04\\_Forsyth.doc](http://www.statistics.gov.uk/events/-q2006/downloads/T04_Forsyth.doc))
- Fowler, F.J. (1992): How Unclear Terms Affect Survey Data. *Public Opinion Quarterly*. Vol. 56, No. 2.
- Fowler, F.J. (1993): *Survey Research Methods*. Thousand Oaks: Sage Publications.
- Fowler, F.J. (1995)\*: *Improving survey questions: design and evaluation*. Thousand Oaks, Sage Publications.
- Fowler, F.J. (2002)\*: *Survey Research Methods*. Thousand Oaks, Sage Publications.
- Fowler, F.J. & Mangione, T.W. (1990): *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park: Sage Publications.
- Fowler, F.J. & Mangione, T.W. (1991): *Standardized Survey Interviewing*. Newbury Park, California.
- Frey, F. W. (1970)\*: Cross-cultural survey research in political science. I: Holt, R.T. & Turner, J.E. (Eds.), *The methodology of comparative research*. New York: Free Press.
- Frey, J.H. & Oishi, S.M. (1995): *How to Conduct Interviews by Telephone and in Person*. Thousand Oaks: Sage Publications.
- Geer, J.G. (1991): Do Open-Ended Questions Measure "Salient" Issues? *Public Opinion Quarterly*. Vol. 55.
- Goerman, P.L. (2006)\*: Adapting Cognitive Interview Techniques for the Use in Pretesting Spanish Language Survey Instruments. Washing-

- ton: U.S. Census Bureau (<http://www.census.gov/srd/papers/-pdf/rsm2006-03.pdf>).
- Greenfield, P.M. (1997)\*: You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist*. Vol. 52: 1115-1124.
- Griching, W.L. (1986): Amount and Direction of Bias in Survey Interviewing. *Australian Psychologist*. Vol. 21, No. 1.
- Groves, R.M. et al. (eds.) (1988): Telephone Survey Methodology. New York: Wiley.
- Groves, R.M. (1989): Survey Errors and Survey Costs. New York: Wiley.
- Groves, R.M. et al. (eds.) (2001): Survey Nonresponse. New York: Wiley.
- Grunert, S.C. & Muller, T.E. (1996)\*: Measuring Values in International Settings. *Journal of International Consumer Marketing*. Vol. 8: 169-185.
- Hakim, C. (1987): Research Design. Boston.
- Hambleton, R.K. & Patsula, L. (1998)\*: Adapting tests for use in multiple languages and cultures. *Social Indicators Research*. Vol. 45: 153-171.
- Hansen, E.J. & Andersen, B.H. (2000): Et sociologisk værktøj: Introduktion til den kvantitative metode. København: Hans Reitzels Forlag.
- Hantrais, L. & Mangen, S. (eds.) (1996)\*: Cross-national research methods in the social sciences. London: Pinter Publishers.
- Haraldsen, G. (1999): Spørreskemametodikk etter kokebokmetoden. Oslo: Gyldendal.
- Harboe, T. (2001): Indføring i samfundsvidenskabelig metode. København: Samfundslitteratur.
- Harkness, J.A. (1996)\*: Cognitive approaches to survey translation. International Sociological Association Conference on Social Science Methodology. University of Essex.
- Harkness, J.A. (1999)\*: In pursuit of quality: issues for cross-national survey research. *International Journal of Social Research Methodology*. Vol. 2, No. 2: 125-140.
- Harkness, J.A. (ed.) (2005)\*: Conducting Cross-National and Cross-Cultural Surveys. Papers from the 2005 Meeting of the International Workshop on Comparative Survey Design and Implementation. Mannheim: ZUMA.
- Harkness, J.A. et al. (eds.) (2003)\*: Cross-Cultural Survey Methods. New York: Wiley-Interscience.

- Harkness, J.A. et al. (2004)\*: Survey Questionnaire Translation and Assessment. I: Presser, S. et al. (eds.), *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.
- Harkness, J.A. & Schoua-Glusberg, A. (1998)\*: Questionnaires in Translation. I: ZUMA-Nachrichten Spezial No. 3. Cross-Cultural Survey Equivalence. Mannheim: ZUMA ([http://www.gesis.org/Publikationen/Zeitschriften/ZUMA\\_Nachrichten\\_spezial/documents/znspezial3/znspez3\\_04\\_Harkness\\_Glusberg.pdf](http://www.gesis.org/Publikationen/Zeitschriften/ZUMA_Nachrichten_spezial/documents/znspezial3/znspez3_04_Harkness_Glusberg.pdf)).
- Harpaz, I. (2000)\*: The Essence of Performing Meaningful Comparative International Survey Research. Center for the Study of Organizations & Human Resource Management. University of Haifa ([http://organizations.haifa.ac.il/html/html\\_eng/book%20-%20harpaz2.doc](http://organizations.haifa.ac.il/html/html_eng/book%20-%20harpaz2.doc)).
- Harrison, D.A. & McLaughlin, M.E. (1993): Cognitive Processes in Self-Report Responses. *Journal of Applied Psychology*. Vol. 78/1.
- Harzing, A-W. (2005)\*: Does the Use of English-language Questionnaires in Cross-national Research Obscure National Differences? *International Journal of Cross Cultural Management*. Vol. 5, No. 2: 213-224.
- Harzing, A-W. (2006)\*: Response Styles in Cross-national Survey Research. *International Journal of Cross Cultural Management*. Vol. 6, No. 2: 243-266.
- Hasher, L. & Griffin, M. (1978): Reconstructive and Reproductive Processes in Memory. *Journal of Experimental Psychology*. Vol. 4, No. 4.
- Heath, A. et al. (2005)\*: The globalization of public opinion research. *Annual Review of Political Science*. Vol. 8: 297-333.
- Hellevik, O. (1993a): *Forskningsmetode i sosiologi og statsvitenskap*. Oslo: Universitetsforlaget.
- Hellevik, O. (1993b): *Sosiologisk metode*. Oslo: Universitetsforlaget.
- Hellevik, O. (2002): *Forskningsmetode i sosiologi og statsvitenskap*. Oslo: Universitetsforlaget.
- Hendershot, G.E. & Crews, J.E. (2006)\*: Toward International Comparability of Survey Statistics on Visual Impairment. *Journal of Visual Impairment and Blindness*. No. 1: 11-25.
- Henry, G.T. (1990): *Practical Sampling*. Newbury Park: Sage Publications.

- Herdman, M. et al. (1997)\*: 'Equivalence' and the translation and adaptation of health-related quality of life questionnaires. *Quality of Life Research*. Vol. 6, No. 3.
- Hines, A.M. (1993)\*: Linking Qualitative and Quantitative Methods in Cross-Cultural Survey Research. *American Journal of Community Psychology*. Vol. 21: 729-746.
- Hippler, H-J. et al. (eds.) (1987): Social Information Processing and Survey Methodology. New York: Springer Verlag.
- Hippler, H-J. & Schwarz, N. (1986): Not Forbidding isn't Allowing. *Public Opinion Quarterly*. Vol. 50, No. 1.
- Hoffmeyer-Zlotnik, J.H.P. (2004)\*: Data Harmonization. Luxembourg: NESSIE (<http://www.nessie-essex.co.uk/>)
- Hoffmeyer-Zlotnik, J.H.P. & Harkness, J.A. (eds.) (2005)\*: Methodological Aspects in Cross-National Research. ZUMA-Nachrichten Spezial, Vol 11. Mannheim: ZUMA 2005.
- Hoffmeyer-Zlotnik, J.H.P. & Wolf, C. (2005)\*: Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables. Berlin: Springer.
- Hogarth, R.M. (ed.) (1982): Question Framing and Response Consistency. San Francisco, Washington & London.
- Hogg, M.A. & Vaughan, G.M. (1995): Social Psychology. London.
- Holme, I.M. & Solvang, B.K. (1996): Metodevalg og metodebruk. Oslo: Tano.
- Hosmer, D.W. & Lemeshov, S. (1989): Applied Logistic Regression. New York.
- Houtkoop-Steenstra, H. (2000): Interaction and the Standardized Survey Interview. Cambridge University Press.
- Hudler, M. & Richter, R. (2000)\*: Source-book about Questions on Social Reporting in Cross-national and Cross-sectional Surveys. Wien: EU Reporting Working Paper No. 13 (<http://members.aon.at/plg/eu-reporting/sourcebook.pdf>).
- Hudler, M. & Richter, R. (2002)\*: Cross-national Comparison of the Quality of Life in Europe. *Social Indicators Research*, 58: 217-228.

- Hui, C.H. & Triandis, H.C. (1989)\*: Effects of Culture and Response Format on Extreme Response Style. *Journal of Cross-Cultural Psychology*. Vol. 20: 296-309.
- Hyman, H.H. (1955): Survey Design and Analysis. Illinois.
- Hyman, H.H. (1991): Taking Society's Measure: A Personal History of Survey Research. New York: Russel Sage Publications.
- Ilstad, S. (1989): Surveymetoden: En veiledning i utvalgsundersøkelser. Trondheim: Tapir.
- Inglehart, R. & Welzel, C. (2004)\*: What Insights can Multi-Country Surveys Provide about People and Societies? American Political Science Association: *Comparative Politics Newsletter* ([http://www.worldvalues-survey.org/Upload/5\\_APSA-CPChris.doc](http://www.worldvalues-survey.org/Upload/5_APSA-CPChris.doc)).
- Iyengar, S. (1983)\*: Assessing linguistic equivalence in multilingual surveys. I: Bulmer, M. and Warwick, D.P. (eds.), Social research in developing countries. New York: Wiley.
- Jabine, T.B. et al. (eds.) (1984): Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. Washington: National Research Council.
- Jackson, C.J. & Furnham, A. (2000): Designing and analysing questionnaires and surveys. London: Whurr.
- Jackson, W. (1988): Survey Methods: Rules for Survey Design and Analysis. Prentice-Hall.
- Javeline, D. (1999)\*: Response Effects in Polite Cultures. A Test of Acquiescence in Kazakhstan. *Public Opinion Quarterly*, 63: 1-28.
- Jiang, S. et al. (2006)\*: Issues in Cross-cultural Criminal Justice Survey Research. *Internet Journal of Criminology* ([www.internetjournalofcriminology.com](http://www.internetjournalofcriminology.com)).
- Ji, L.J. et al. (2000)\*: Culture, autobiographical memory, and behavioral frequency reports: Measurement issues in cross-cultural studies. *Personality and Social Psychology Bulletin*. Vol. 26: 585-593.
- Jobe, J.B. & Mingay, D.J. (1989): Cognition and Survey Measurement. *Applied Cognitive Psychology*. Vol. 5, no. 3.
- Jobe, J. & Mingay, D. (1990)\*: Cognitive laboratory approach to designing questionnaires for surveys of the elderly. *Public Health Reports*. No. 5: 518-23.



- Jobe, J.B. & Mingay, D.J. (1991): Cognition and Survey Measurement: History and Overview. *Applied Cognitive Psychology*. Vol. 5, No. 3.
- Jobe, J.B. et al. (1993): Contributions of Survey Research to the Understanding of Memory. *Applied Cognitive Psychology*. Vol. 7.
- Johannessen, A. & Tufte, P.A. (2002): Introduksjon til samfunnsvitenskapelig metode. Oslo: Abstrakt.
- Johnson, T. P. et al. (1997)\*: Social cognition and responses to survey questions among culturally diverse populations. I: Lyberg, L. et al. (eds.), Survey measurement and process quality. New York: John Wiley & Sons.
- Johnson, T. (1998)\*: Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA-Nachrichten Spezial*, 3: 1–40.
- Johnson, T. (2006)\*: Methods and Frameworks for Cross-cultural Measurement. Qualitative Issues and Approaches. *Medical Care*. Vol. 44, Suppl. 3: 17-20.
- Johnson, T. et al. (2001)\*: Respondent Cultural Orientations and Survey Participation. 2000 Proceedings of the Section on Survey Research Methods. Alexandria: American Statistical Association ([http://www.amstat.org/sections/SRMS/Proceedings/papers/2000-\\_161.pdf](http://www.amstat.org/sections/SRMS/Proceedings/papers/2000-_161.pdf)).
- Johnson, T. et al. (2005)\*: The relation between culture and response styles: Evidence from 19 Countries. *Journal of Cross-Cultural Psychology*, 36: 264-277.
- Jolliffe, F.R. (1986): Survey Design and Analysis. Chichester: Ellis Horwood.
- Jowell, R. (1998)\*: How comparative is comparative research? *American Behavioral Scientist*. No. 2: 168-177.
- Judd, C.M. et al. (1991): Research Methods in Social Relations. Philadelphia: Hartcourt Brace.
- Jylhä, M. et al. (1998)\*: Is self-rated health comparable across cultures and genders? *Journals of Gerontology*. Series B. Vol. 53, Issue 3, 144-152.
- Jæger, M.M. (2006): Welfare Regimes and Attitudes Towards Redistribution: The Regime Hypothesis Revisited. *European Sociological Review*. Vol. 22, No. 2:157-170.
- Kahn, R. & Canell, C.F. (1957): The Dynamics of Interviewing. New York.

- Kalton, G. & Schuman, H. (1982): The Effect of the Question on Survey Responses: A Review. *Journal of the Royal Statistical Society. Series A*, Vol. 145, No. 1.
- Katan, D. (1999)\*: *Translating Cultures*. Manchester: St. Jerome Publ.
- Kent, R.A. (2001): *Data Construction and Data Analysis for Survey Research*. Palgrave Macmillan.
- Kerlinger, F.N. (1986): *Foundations of Behavioral Research*. New York.
- King, G. et al. (2003)\*: Enhancing the Validity and Cross-Cultural Comparability of Survey Research, *American Political Science Review*, Vol. 97, No. 4: 567-584.
- Knäuper, B. et al (1995): The Quality of Survey Data as Affected by Question Difficulty and Respondents' Cognitive Capacity. 50th Conference of the American Association for Public Opinion Research. Michigan: Institute for Social Research.
- Knäuper, B. et al. (1997)\*: Question difficulty and respondents' cognitive ability: The effect of data quality. *Journal of Official Statistics*, 13: 181-99.
- Krebs, D. & Schmidt, P. (eds.) (1993): *New Directions in Attitude Measurement*. Berlin: Walter de Gruyter.
- Kreiner, S. (1999) *Statistisk problemløsning: Præmisser, teknik og analyse*. København: Jurist- og Økonomforbundets Forlag.
- Krosnick, J.A. (1991): Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*. Vol. 5, No. 2.
- Krosnick, J.A. (1999): Survey Research. *Annual Review of Psychology*, Annual. (<http://arjournals.annualreviews.org/doi/pdf/10.1146/annurev.psych.50.1.537>).
- Krosnick, J.A. (1999)\*: Survey Research. *Annual Review of Psychology*. Vol. 50 (1999), 537-67.
- Krosnick, J.A. & Alwin, D.F. (1987): An Evaluation of a Cognitive Theory of Response Order Effects in Survey Measurement. *Public Opinion Quarterly*. Vol. 51, No. 2.
- Kuechler, M. (1987)\*: The utility of surveys for cross-national research. *Social Science Research*, 16: 229-244.

- Kuechler, M. (1998)\*: The survey method - An indispensable tool for social science research everywhere? *American Behavioral Scientist*. Vol. 42: 178-200.
- Kumar, R. (1999): *Research Methodology: A Step-by-Step Guide*. Thousand Oaks: Sage Publications.
- Kvale, S. (1997): *InterView. En introduktion til det kvalitative forskningsinterview*. Kbh.: Hans Reitzels Forlag.
- Körmendi, E. (1986): *Datakvalitet ved telefoninterview*. København: Socialforskningsinstituttet.
- Labaw, P.J. (1980): *Advanced Questionnaire Design*. Cambridge.
- Lalwani, A.K. et al. (2006)\*: What Is the Relation between Cultural Orientation and Socially Desirable Responding? *Journal of Personality and Social Psychology*. No. 1: 165-178.
- Landman, T. (2003)\*: *Issues and Methods in Comparative Politics*. London: Routledge.
- Langlet, P. & Wärneryd, B. (1983): *Att fråga. Om frågekonstruktion vid intervju och enkätundersökningar*. Stockholm: Liber.
- Lantz, A. (1993): *Intervjumetodik: Den professionellt genomförda intervjun*. Lund: Studentlitteratur.
- Larsen, S.F. (1983): Erindringens natur og historie. *Psyke & Logos*. Vol. 4, No. 2.
- Larsen, S.F. et al. (1994): *Time in Autobiographical Memory*. University of Aarhus.
- Leahey, T.H. (1992): *A History of Psychology. Main Currents in Psychological Thought*. New Jersey.
- Leinhardt, S. (ed.) (1981): *Sociological Methodology*. San Francisco.
- Lessler, J.T. (1987): *Use of Laboratory Methods and Cognitive Science for the Design and Testing of Questionnaires*. Statistics Sweden. Stockholm.
- Lessler, J.T. et al. (1989): *Questionnaire Design in the Cognitive Research Laboratory. Vital and Health Statistics. Series 6, No. 1*.
- Lessler, J.T. & Kalsbeek, W.D. (1992): *Nonsampling Error in Surveys*. New York: Wiley.

- Lessler, J.T. & Sirken, M.G. (1985): Laboratory-based Research on the Cognitive Aspects of Survey Methodology. *Milbank Memorial Fund Quarterly*. Vol. 63, No. 3.
- Levine, Robert V. (1988)\*: The pace of life across cultures. I: McGrath, J.E. (ed.), *The Social Psychology of Time*. Newbury Park: Sage Publications.
- Lewy, P.S. & Lemeshow, S. (1999): *Sampling of Populations: Methods and Applications*. New York: Wiley.
- Litwin, M.S. (1995): *How to Measure Survey Reliability and Validity*. Thousand Oaks: Sage Publications.
- Loftus, E.F. et al. (1985): Cognitive Psychology Meets the National Survey. *American Psychologist*. Vol. 40, No. 2.
- Lonner, W. J. (1990)\*: An overview of cross-cultural testing and assessment. I: Brislin, R.W. (Ed.), *Applied cross-cultural psychology* (pp. 56-76). Newbury Park, CA: Sage.
- Lund, T. et al. (red.) (2002): *Innføring i forskningsmetodologi*. Oslo: Unipup.
- Lyberg, L. et al. (Eds.) (1997)\*: *Survey measurement and process quality*. New York: Wiley.
- Lynn, P. (2003)\*: Developing quality standards for cross-national survey research. *International Journal of Social Research Methodology*. Vol. 6, No. 4: 323-336.
- MacIntosh, R. (1998)\*: Global attitude measurement: An assessment of the World Values Survey postmaterialism scale, *American Sociological Review*, 63, 452-464.
- Manderbacka, K. (1998)\*: How do respondents understand survey questions on ill-health? *European Journal of Public Health* 8 (4):319-324, Dec. 1998.
- Marsh, C. (1982): *The Survey Method: The Contribution of Surveys to Sociological Explanation*. London.
- Mathiowitz, N. & Groves, R.M. (1985): The Effects of Respondent Rules on Health Survey Reports. *American Journal of Public Health*. Vol. 75.
- Maynard, D.W. et al. (eds.) (2001): *Standardization and tacit knowledge. Interaction and practice in the survey interview*. New York: Wiley.
- McFarland, S.G. (1981): Effects of Question Order on Survey Responses. *Public Opinion Quarterly*. Vol.45, No. 2.

- McKay, R.B. et al. (1996)\*: Translating Survey Questionnaires: Lessons Learned. *New Directions for Evaluation*, 70: 93-105.
- McLaughlin, M.E. (1999): Controlling Method Effects in Self-Report Instruments. Department of Management University of Texas at Arlington ([www.aom.pace.edu/rmd/1999\\_RMD\\_Forum\\_Method\\_Effects\\_in\\_Self-Reports.htm](http://www.aom.pace.edu/rmd/1999_RMD_Forum_Method_Effects_in_Self-Reports.htm)).
- Menard, S. (1995): Applied Logistic Regression Analysis. Series: Quantitative Applications in the Social Sciences. Thousand Oaks: Sage Publications.
- Merton, R.K. et al. (1990): The Focused Interview. A Manual of Problems and Procedures. New York.
- Middleton, K.L. & Jones, J.L. (2000)\*: Socially desirable response sets: The impact of country culture. *Psychology & Marketing*, 17: 149-163.
- Miller, K. (2003)\*: Conducting Cognitive Interviews to Understand Question Response Limitations. *American Journal of Health Behavior*. Vol. 27, No. 3: 264-272.
- Miller, K. (2004)\*: Implications of Socio-cultural Factors in the Question Response Process. ZUMA-Nachrichten Spezial Band.
- Mills, M. et al. (2006)\*: Comparative Research. Persistent Problems and Promising Solutions. *International Sociology*. Vol. 21: 619-631.
- Mingay, D.J. & Greenwell, M.T. (1989): Memory Bias and Response-Order Effects. *Journal of Official Statistics*. Vol. 5, No.3.
- Mishler, E.G. (1979): Meaning in Context: Is There Any Other Kind? *Harvard Educational Review*. Vol. 49, No.1.
- Mohler, P. (2005)\*: Challenges of Comparative Surveys. Mannheim: ZUMA ([http://www.dji.de/cgi-bin/inklunde.php?inklunde=9\\_themen/thema-0508/outsideview.htm](http://www.dji.de/cgi-bin/inklunde.php?inklunde=9_themen/thema-0508/outsideview.htm))
- Mohler, P. et al. (1998)\*: Respondents' Ratings of Expressions from Response Scales. ZUMA Spezial ([http://www.gesis.org/Publikationen/Zeitschriften/ZUMA\\_Nachrichten\\_spezial/documents/znspezial3/znspez3\\_07\\_Mohler\\_Smith\\_Harkness.pdf](http://www.gesis.org/Publikationen/Zeitschriften/ZUMA_Nachrichten_spezial/documents/znspezial3/znspez3_07_Mohler_Smith_Harkness.pdf)).
- Moors, G. (2004)\*: Facts and Artefacts in the Comparison of Attitudes among Ethnic Minorities. *European Sociological Review*. Vol. 20, No. 4: 303-320.
- Mordal, T.L. (1989a): Som man spør, får man svar. Oslo: Tano.

- Mordal, T.L. (1989b): Vi spør - hvor mange og hvem svarer? Oslo: Statens Institutt for Forbruksforskning.
- Mordal, T.L. (2000): Som man spør, får man svar. Oslo: Tano.
- Morton-Williams, J. (1993): Interviewer approaches. Alderhot: Dartmouth.
- Morton-Williams, J. & Sykes, W. (1984): The Use of Interaction Coding and Follow-up Interviews to Investigate Comprehension of Survey Questions. *Journal of the Market Research Society*. Vol. 26, No. 2.
- Moser, C.A. & Kalton, G. (1993): Survey methods in social investigations. Alderhot: Dartmouth.
- Mullen, M.R. (1995)\*: Diagnosing Measurement Equivalence in Cross-National Research. *Journal of International Business Studies*, Vol. 26.
- Nachmias, D. & Nachmias, C. (1987): Research Methods in the Social Sciences. New York.
- Nardi, P.M. (2003): Doing survey research. Boston: Allyn and Bacon.
- Nassar-McMillan, S.C. & Borders, L.D. (2002): Use of Focus Groups in Survey Item Development. *The Qualitative Report*. Vol. 7, nr. 1 ([www.nova.edu/ssss/QR/QR7-1/nassar.html](http://www.nova.edu/ssss/QR/QR7-1/nassar.html)).
- Neuman, W.L. (2000): Social Research Methods: Qualitative and Quantitative Approaches. Allyn and Bacon.
- Nicoletti, C. & Peracchi, F. (2005)\*: Survey response and survey characteristics (ECHP). *Journal of the Royal Society*. 168:763-781, part 4, 2005.
- Nielsen, P. & Nicolajsen, A.J. (1988): Vejledning i spørgeskemaundersøgelse med anvendelse af statistikprogrammet SAS. København.
- Nolin, M.J. & Chandler, K. (1996): Use of cognitive laboratories and recorded interviews in the National Household Education Survey. Washington: National Center for Education Statistics.
- Nolinske, T. (1998): Survey Research and Measurement Error. Washington ([www.oandp.org/jpo/library/1995\\_02\\_068.asp](http://www.oandp.org/jpo/library/1995_02_068.asp)).
- Okazaki, S. & Sue, S. (1995)\*: Methodological Issues in Assessment Research with Ethnic Minorities. *Psychological Assessment*, 7: 367-375.
- Oksenberg, L. et al. (1991): New Strategies for Pretesting Survey Questions. *Journal of Official Statistics*. 1991/7:349ff.
- Olsen, H. (1995): Tallenes talende tavshed (I). København: Socialforskningsinstituttet.

- Olsen, H. (1997): Tal taler ikke uden ord. Et metodeeksperiment om danske respondenteres sprogsensitivitet i politologiske og sociologiske surveyundersøgelser. *Politica*. 3/1997.
- Olsen, H. (1998): Tallenes talende tavshed (II). Måleproblemer i surveyundersøgelser. København: Akademisk Forlag.
- Olsen, H. (2000): Holdninger til handicappede. København: Socialforskningsinstituttet.
- Olsen, H. (2001a): Sprogforståelse og hukommelse i danske surveyundersøgelser (bd. I-II). København: Socialforskningsinstituttet.
- Olsen, H. (2001b): ”Hvor enig eller uenig er du?”. Holdningsspørgsmål og holdningsmålinger i danske samfundsvidenskabelige surveyundersøgelser. *Nordiske Udkast*, årg. 29, nr. 2.
- Olsen, H. (2002a): Kvalitative kvaler. Kvalitative metoder og danske kvalitative interviewundersøgelser kvalitets-kvalitet. København: Akademisk Forlag.
- Olsen, H. (2002b): The Silence of Numbers. A Split-sample Experiment concerning Respondent’s Linguistic Sensitivity in Danish Survey Studies. *International Journal of Social Research Methodology*, 4/2002.
- Olsen, H. (2004): *De plukkes ikke som blomster på en mark ...* Tendenser i kvantitativ metodelitteratur vedrørende konstruktion og kvalitets-sikring af surveydata. *Dansk Sociologi* 2004/2 .
- Olsen, H. (2005a): Tværvideenskabelig spørgeskemakonstruktion. *Sociologisk forskning* (svensk), 1/2005.
- Olsen, H. (2005): Fra spørgsmål til svar. København: Akademisk Forlag.
- Olsson, H. & Sörensen, S. (2002): Forskningsprocessen: Kvalitative og kvantitative aspekter. Oslo: Gyldendal.
- Oppenheim, A.N. (1992): Questionnaire Design, Interviewing, and Attitude Measurement. London: Pinter Publishers.
- Osgood, C.E. et al. (1971): The Measurement of Meanings. University of Illinois: Illini Books.
- O’Shea, R. et al. (2005)\*: Comparative Attitudinal Research in Europe. London: National Centre for Social Research ([http://naticent02.uuhost.uk.uu.net/questionnaire/questionnaire\\_development/comparative\\_attitudinal\\_research.pdf](http://naticent02.uuhost.uk.uu.net/questionnaire/questionnaire_development/comparative_attitudinal_research.pdf)).

- Oskamp, S. (1977): Attitudes and Opinions. Englewood Cliffs: Prentice-Hall Inc.
- Ottati, V.C. et al. (1989): Cognitive and Affective Bases of Opinion Survey Responses. *Journal of Personality and Social Psychology*. Vol. 57.
- Pan, Y. (2003)\*: The role of sociolinguistics in the development and conduct of federal surveys. US Census Bureau (<http://www.fcs.gov/03papers/Panfinal.pdf>).
- Pareek, U. & Rao, T.V. (1980)\*: Cross cultural survey and interviewing. I: Triandis, H.C. and Berry, J.W. (eds.), *Handbook of cross-cultural psychology*, 2, Boston: Allyn and Bacon.
- Pasick, R. et al. (1996)\*: Problems and progress in translation of health survey questions. *Health Education Quarterly*, 23: 28-40.
- Patel et al., J. (2005)\*: Cultural Response Bias. UK: Customer Equity Comp. (<http://www.samra.co.za/2006papers/8.%20Cultural%20Response%20Bias.doc>).
- Pauwels, L. & Pleysier, S. (2005)\*: Assessing Cross-Cultural Validity of Fear of Crime Measures through Comparisons between Linguistic Communities in Belgium. *European Journal of Criminology*, Vol. 2, No. 2: 139-159.
- Payne, S.L. (1951): The Art of Asking Questions. New Jersey.
- Peterson, R.A. (2000): Constructing Effective Questionnaires. Thousand Oaks: Sage Publications.
- Philips, D.L. (1971): Knowledge from what? Chicago.
- Philips, L. R. et al. (1994)\*: Strategies for achieving cultural equivalence. *Research in Nursing & Health*, 17: 149-154.
- Poe, G.S. et al. (1988): "Don't Know" Boxes in Factual Questions. *Public Opinion Quarterly*. Vol. 55.
- Poortinga, Y.H. (1989)\*: Equivalence of cross-cultural data: An overview. *International Journal of Psychology*, 24: 737-756.
- Pratkanis, A.R. et al. (1989): Attitude Structure and Function. Hillsdale, New Jersey: Lawrence Erlbaum Ass., Publ.
- Presser, S. (1990): Measurement Issues in the Study of Social Change. *Social Forces*. Vol. 68, No. 3.
- Presser, S. (1990a): Can Context Changes Reduce Vote Overreporting? *Public Opinion Quarterly*. Vol. 54, No. 4.



- Presser, S. & Blair, J. (1994): Do Different Methods Produce Different Results? *Sociological Methodology*, pp. 73-104.
- Presser, S. & Schuman, H. (1980): The Measure of a Middle Position in Attitude Surveys. *Public Opinion Quarterly*. Vol. 44.
- Presser, S. et al. (2004)\*: Methods for Testing and Evaluating Survey Questions. *Public Opinion Quarterly*. Vol. 68, No. 1: 109-130.
- Presser, S. et al. (2004a)\*: Methods for Testing and Evaluating Survey Questionnaires. Wiley.
- Przeworski, A. & Teune, H. (1970)\*: The Logic of Comparative Social Inquiry. New York: Wiley.
- Punch, K.F. (1998): Introduction to Social Research: Quantitative and Qualitative Approaches. London: Sage Publications.
- Raitasalo, K. et al. (2005)\*: Retrieval strategies and cultural differences in answering survey questions on drinking. *Addiction Research & Theory*. Vol. 13, No. 4: 359-372.
- Rajecki, D.W. (1990): Attitudes. Sunderland: Sinauer Ass.
- Rasinski, K. (1989): The Effect of Question Wording on Public Support for Government Spending. *Public Opinion Quarterly*, nr. 53.
- Rasmussen, E.S. & Østergaard, P. (2002): Samfundsvidenskabelige metoder. Odense: Odense Universitetsforlag.
- Rea, L.M. & Parker, R.A. (1997): Designing and Conducting Survey Research. San Francisco: Jossey-Bass Publications.
- Redmond, M. (2003)\*: Cultural and ethical challenges in cross-national research. *International Journal of Qualitative Methods*. Vol. 2, No. 4: 1-21.
- Reis, H.T. & Judd, C.M. (eds.) (2000): Handbook of Research methods in social and personality psychology. Cambridge: Cambridge University Press.
- Riis, O. (2001): Metoder på tværs: Om forudsætningerne for sociologisk metodekombination. København: Jurist- og Økonomforbundets Forlag.
- Ringdal, K. (2001): Enhet og mangfold: Samfunnsvitenskapelig forskning og kvantitativ metode. Bergen.
- Robert, C. et al. (2006)\*: An Empirical Analysis of Measurement Equivalence with the INDCOL Measure of Individualism and Collectivism.

- Personnel Psychology*. No. 59: 65-99 (<http://www.blackwellsynergy.com/doi/pdf/10.1111/j.1744-6570.2006.00804.x>).
- Rode, N. (2005)\*: Translation of Measurement Instruments and their Reliability. *Metodoloski zvezki*. Vol. 2, No.1: 15-26 (<http://mrvar.fdv.unilj.si/pub/mz/mz2.1/rode.pdf>).
- Rokkan, S. (1993)\*: Cross-cultural, cross-societal and cross-national research. Definitions and implementations of different forms of “macro level” comparisons. *Historical Social Research*, 18, 6-54.
- Ronkainen, S. (1998): Impossibility of Reality and Possibility of the Survey as a Construction. *Nordiske Udkast*, nr. 2/1998, pp. 51-61.
- Ross, C. E., & Mirowsky, J. (1984)\*: Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *Journal of Health and Social Behavior*, 25, 189-197.
- Rossi, P.H. et al. (1983): *Handbook of Survey Research*. London.
- Rowney, J. & Taras, V. (2006)\*: Half a century of measuring culture: Approaches, challenges, limitations and suggestions. Alberta: University of Calgary ([http://www.ucalgary.ca/~taras/\\_private/measuring\\_culture.doc](http://www.ucalgary.ca/~taras/_private/measuring_culture.doc)).
- Salant, P. & Dillman, D.A. (1994): *How to Conduct your own Survey*. New York: Wiley.
- Salzberger, T. (2000)\*: Intercultural Construct Validity in Emic and Etic Research. Proceedings of the 2000 Multicultural Marketing Conference: Wien. ([http://www2.wu-wien.ac.at/marketing/user/salzberger/research/amsmulti-cultural2000\\_salzberger.pdf](http://www2.wu-wien.ac.at/marketing/user/salzberger/research/amsmulti-cultural2000_salzberger.pdf)).
- Sanders, D. (1994)\*: Methodological considerations in comparative cross-national research. *International Social Science Journal*. Vol. 46, No. 4: 513-521.
- Sapsford, R. (1999): *Survey Research*. London: Sage Publ.
- Saris, W.E. (1998)\*: The Effects of Measurement Error in Cross-Cultural Research. *Nachrichten Spezial, Cross-Cultural Survey Equivalence 3*. ([http://www.gesis.org/Publikationen/Zeitschriften/ZUMA\\_Nachrichten\\_spezial/documents/znspezial3/znspez3\\_03\\_SARIS.pdf](http://www.gesis.org/Publikationen/Zeitschriften/ZUMA_Nachrichten_spezial/documents/znspezial3/znspez3_03_SARIS.pdf)).
- Sasaki, M. (1995)\*: Research Design of Cross-National Attitude Surveys. *Behaviormetrica*, 22: 99-114.

- Schaeffer, N.C. & Presser, S. (2003)\*: The Science of Asking Questions. *Annual Review of Sociology* 29: 65–88.
- Schaffer, B.S. & Riordan, C.M. (2003)\*: A review of cross-cultural methodologies for organizational research: A best-practices approach, *Organizational Research Methods*, 6: 169-215.
- Scheuch, E.K. (1989)\*: Theoretical Implications of Comparative Survey Research. *International Sociology*. Vol. 4: 147-167.
- Scheuch, E. K. (1993)\*: The cross-cultural use of sample surveys: Problems of comparability. *Historical Social Research*, 18: 104–38.
- Schuman, H. & Converse, J. (1971): The Effect of Black and White Interviewers on Black Responses. *Public Opinion Quarterly*. Vol. 3, p. 44ff.
- Schuman, H. & Presser, S. (1977): Question Wording as an Independent Variable in Survey Analysis. *Sociological Methods and Research*. Vol. 6, No. 2.
- Schuman, H. & Presser, S. (1981): Questions and Answers in Attitude Surveys. New York.
- Schuman, H. & Presser, S. (1996): Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. Thousand Oaks: Sage Publications.
- Schuman, H. et al. (1981): Context Effects on Survey Responses to Questions about Abortion. *Public Opinion Quarterly*. Vol. 45.
- Schuman, H. et al. (1983): Context and Contiguity in Survey Questionnaires. *Public Opinion Quarterly*. Vol. 47, No.1.
- Schwarz, N. (1999): Self-reports: How the questions shape the answers. *American Psychologist*, 54:93-105.
- Schwarz, N. et al. (1991): The Impact of Administration Mode on Response Effects in Survey Measurement. *Applied Cognitive Psychology*. Vol. 5, No. 3.
- Schwarz, N. & Hippler, H.J. (1995)\*: The Numeric Values of Rating Scales. *International Journal of Public Opinion Research*, 7: 72-74.
- Schwarz, N. & Sudman, S. (eds.) (1992): Context Effects in Social and Psychological Research. New York.
- Schwarz, N. & Sudman, S. (eds.) (1996)\*: Answering Questions. Methodology for determining cognitive and communicative processes in survey research. New York: Jossey-Bass Publications.

- Shenkar, O. & von Glinow, M.A. (1994)\*: Paradoxes of organizational theory and research. *Management Science*. Vol. 40, No. 1: 56-71.
- Singer, E. & Presser, S. (eds.) (1989): Survey Research Methods: A Reader. Chicago: University of Chicago Press.
- Singh, N.N. et al. (2000)\*: Semantic Equivalence of Assessment Instruments across Cultures. *Journal of Child and Family Studies*. Vol. 9, No. 2: 123-134.
- Singleton, R. et al. (1988): Approaches to Social Research. New York.
- Sireci, S.G. et al. (2003)\*: Evaluating the Equivalence of an Employee Attitude Survey across Languages, Cultures, and Administration Formats. *International Journal of Testing*, Vol. 3, No. 2: 129-150.
- Sirken, M.G. et al. (ed.) (1999): Cognition and Survey Research. New York: Wiley.
- Sjöström, S. (2001)\*: Semantisk förändring. Hur ord får nya betydelser. Stockholm: Studentlitteratur.
- Skaksen, J.R. et al. (2006)\*: Holdninger til immigration: Er danskere specielle? København: Centre for Economic and Business Research.
- Skinner, J.K. (2001)\*: Acculturation: Measures of Ethnic Accommodation to the Dominant American Culture. *Journal of Mental Health and Aging*. Vol. 7, no. 1:41-51
- Skinner, C.J. (1989): Analysis of Complex Surveys. Chichester: Wiley.
- Small, R. et al. (1999)\*: Cross-cultural research: Trying to do it better 2. Enhancing data quality. *Australian and New Zealand Journal of Public Health*, 23: 390-395.
- Smith, A.F. et al. (1991): Retrieval from Memory of Dietary Information. *Applied Cognitive Psychology*. Vol. 5.
- Smith, P.B. & Bond, M.H. (1998)\*: Social Psychology across Cultures. Boston: Allyn & Bacon.
- Smith, T.W. (1987): That Which We Call Welfare by Any Other Name Would Smell Sweeter. *Public Opinion Quarterly*. No. 51.
- Smith, T.W. (2003)\*: Developing Comparable Questions in Cross-National Surveys. I: Harkness, J. et al. (eds.), *Cross-Cultural Survey Methods*. NJ: John Wiley & Sons.
- Smith, T.W. (2004)\*: Crossnational Survey Research: The challenge and the Promise. *ICPSR Bulletin*. Vol. 24, No. 2: 3-12.

- Smith, P.B. (2004a)\*: Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, 35: 50-61.
- Smith, T.W. (2004b)\*: Developing and Evaluating Cross-National Survey Instruments. I: Presser, S. et al., *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.
- Sperber, A.D. et al. (1994)\*: Cross-cultural translation: Methodology and validation. *Journal of Cross-Cultural Psychology* 25: 501-524.
- Sproull, N.L. (1988): *Handbook of Research Methods*. London.
- Stocké, V. (2001)\*: Socially Desirable Response Behavior as Rational Choice. Sonderforschungsbereich 504, No. 1-22. Mannheim: Mannheim Universitet.
- Sudman, S. & Bradburn N.M. (1973): Effects of Time and Memory Factors on Response in Surveys. *Journal of American Statistical Association*. Vol. 68, No. 344.
- Sudman, S. & Bradburn, N.M. (1974): *Response Effects in Surveys: A Review and Synthesis*. Chicago.
- Sudman, S. & Bradburn, N.M. (1982): *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco.
- Sudman, S. & Bradburn, N.M. (1987): *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass.
- Sudman, S. et al. (1996): *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publishers.
- Svensson, G. (1995): Kvantitativ frågekonstruktion. Horred: Adekvat.
- Swanson, G. et al. (2003)\*: Comparing disability survey questions in five countries. *Disability and Rehabilitation* 25 (11-2):665-675.
- Sykes, W. (1985): *Telephone Interviewing on a Survey of Social Attitudes*. London: Social and Community Planning Research.
- Sykes, W. & Morton-Williams, J. (1987): Evaluating Survey Questions. *Journal of Official Statistics*. Vol. 3, no. 2.
- Tanur, J. (1987): Some Cognitive Aspects of Surveys. *Journal of Official Statistics*. Vol. 3, No. 4.
- Tanur, J.M. (ed.) (1992): *Questions about Questions*. New York: Russel Sage Foundation.

- Tanzer, N.K. (2005)\*: Developing Tests for Use in Multiple Languages and Cultures. I: Hambleton, R.K. et al., *Adapting Educational and Psychological Test for Cross-Cultural Assessment*. Hillsdale: Erlbaum
- Temple, B. (1997)\*: Watch your tongue: issues in translation and cross-cultural research. *Sociology*, 31(3): 607-618.
- ten Have, P. (1998): Pragmatic aspects of the production of survey interview data. University of Amsterdam ([www2.fmg.uva.nl/emca/ttc.htm](http://www2.fmg.uva.nl/emca/ttc.htm)).
- Teune, H. (1990)\*: Comparing countries: Lessons learned. I: Oyen, E. (ed.), *Comparative methodology*. London: Sage.
- Therborn, G. (1999)\*: Sociologiens første århundrede – og det næste. *Dansk Sociologi* nr. 4, pp. 19-33.
- Thomas, S.J. (1999): *Designing Surveys that work!* Thousand Oaks: Corwin Press.
- Tourangeau, R. & Rasinski, K.A. (1988): Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin*. Vol. 103.
- Tourangeau, R. et al. (1991): Measuring Happiness in Surveys: A Test of the Subtraction Hypothesis. *Public Opinion Quarterly*. Vol. 55.
- Tourangeau, R. et al. (2000). *The Psychology of Survey Response*. Cambridge University Press.
- Tran, T.V. & Williams, L.F. (1994)\*: Effect of Language of Interview on the Validity and Reliability of Psychological Well-being scales. *Social Work Research*. Vol. 18, No. 1: 17-25.
- Triandis, H.C. (1992)\*: Cross-cultural research in social psychology. I: Granberg, D. & Sarup, F. (eds.), *Social judgment and intergroup relationships*. New York: Springer.
- Triandis, H.C. (1995)\*: *Individualism and Collectivism*. Westview Press.
- Turner, C.E. & Martin, E. (eds.) (1984): *Surveying Subjective Phenomena*. Vol. 1-2. New York.
- Uskul, A.K. & Oyserman, D. (2005)\*: Question Comprehension and Response. University of Michigan (<http://scholar.google.com/scholar?hl=da&lr=&q=%22comprehension+and+response%22+uskul>).
- Usunier, J-C. (1999)\*: The Use of Language in Investigating Conceptual Equivalence in Cross-Cultural Research (<http://marketing.byu.edu/htmlpages/ccrs/proceedings99/usunier.htm>).

- Van Buuren, S. et al. (2005)\*: Improving Comparability of Existing Data by Response Conversion. *Journal of Official Statistics*. Vol. 21, No.1: 53-72.
- Van de Vijver, F.J.R. (1998)\*: Towards a Theory of Bias and Equivalence. Mannheim: ZUMA-Nachrichten Spezial, January 1998: 41-65 ([http://www.gesis.org/publikationen/zeitschriften/ZUMA\\_Nachrichten\\_spezial/documents/znspezial3/znspez3\\_02\\_Vijver.pdf](http://www.gesis.org/publikationen/zeitschriften/ZUMA_Nachrichten_spezial/documents/znspezial3/znspez3_02_Vijver.pdf)).
- Van de Vijver, F.J.R. & Hambleton, R.K. (1996)\*: Translating tests: Some practical guidelines. *European Psychologist* 1: 89-99.
- Van de Vijver, F.J.R. & Leung, K. (1997)\*: Methods and data analysis for cross-cultural research. Thousand Oaks: Sage.
- Van de Vijver, F.J.R. & Leung, K. (2000)\*: Methodological Issues in Psychological Research on Culture. *Journal of Cross-Cultural Psychology*. Vol. 31, No. 1: 33-51
- Van de Vijver, F.J.R. & Poortinga, Y.H. (2002)\*: Structural Equivalence in Multilevel Research. *Journal of Cross-Cultural Psychology*. Vol. 33, No. 2: 141-156.
- Van der Zouwen, J. & Smith, J. (2004)\*: Evaluating Survey Questions by Analyzing Patterns of Behavior Codes and Question-Answer Sequences. I: Presser, S. et al. (eds.): Methods for Testing and Evaluating Survey Questionnaires. New York: Wiley.
- Van Herk, H. (2000)\*: Equivalence in a Cross-National Context. Tilburg: University Press.
- Van Herk, H. et al. (2004)\*: Response styles in rating scales. *Journal of Cross-Cultural Psychology* 35: 346-360.
- van Herk, H. et al. (2005)\*: Equivalence of survey data: relevance for international marketing. *European Journal of Marketing*. Vol. 39, No. 3/4: 351-364.
- Verma, V. (2002)\*: Comparability in International Survey Statistics. Int. Conference on Improving Surveys, Copenhagen ([http://www.icis.dk/ICIS\\_papers/Key-note3\\_0\\_5.pdf](http://www.icis.dk/ICIS_papers/Key-note3_0_5.pdf)).
- Voss, K.E. et al. (1996)\*: An exploration of the comparability of semantic adjectives in three languages: A magnitude estimation approach. *International Marketing Review*. Vol. 13, No. 5: 44-58.

- Walden, G.R. (2003): Survey Research Methodology: An annotated bibliography. Westport: Greenwood Press.
- Warnecke, R. et al. (1997)\*: Improving question wording in surveys of culturally diverse populations. *Annals of Epidemiology*, 7: 334-342.
- Watkins, D. & Cheung, S. (1995)\*: Culture, gender and response bias. *Journal of Cross-Cultural Psychology*, 26: 490-504.
- Weidmer, B. (1994)\*: Issues and Guidelines for Translation in Cross-Cultural Research. California: Santa Monica ([www.amstat.org/sections/SRMS/proceedings/papers/1994\\_215.pdf](http://www.amstat.org/sections/SRMS/proceedings/papers/1994_215.pdf)).
- Weidmer, B. et al. (1999)\*: Translating the Cahps 1.0 Survey Instruments into Spanish. *Medical Care*. No. 37: 89-96.
- Weisberg, H.F. (1996): An introduction to survey research, polling and data analysis. Thousand Oaks: Sage.
- Welkenhuysen-Gybels, J. & Van de Vijver, F.J.R. (2001)\*: Methods for the Evaluation of Construct Equivalence in Studies Involving Many Groups. American Association for Public Research (<http://soc.kuleuven.be/ceso/onderzoek/8/softwaredata/pdf>).
- Wentland, E.J. & Smith, K.W. (1993): Survey Responses: An Evaluation of their Validity. San Diego: Academic Press.
- Wiger, D.E. & Huntley, D.K. (2002): Essentials of interviewing. New York: John Wiley.
- Willis, G.B. et al. (1991): The Use of Verbal Report Methods in the Development and Testing of Survey Questionnaires. *Applied Cognitive Psychology*. Vol. 5, No. 3.
- Willis, G. et al. (2005)\*: The Use of Cognitive Interviewing to Evaluate Translated Survey Questions. Federal Committee on Statistical Methodology ([http://www.fcsm.gov/05papers/Willis\\_Lawrence\\_etal\\_VIIIB.pdf](http://www.fcsm.gov/05papers/Willis_Lawrence_etal_VIIIB.pdf)).
- Wilkinson, D. (2000): The Researcher's Toolkit: The Complete Guide to Practitioner Research. Routledge.
- Wonnacott, T.H. & Wonnacott, R.J. (1990): Introductory Statistics for Business and Economics. New York: John Wiley & Sons.
- Word, C. O. (1992)\*: Cross-cultural methods for survey research in Black urban areas. In Burlew, A.K. et al. (Eds.), African American psy-



chology: Theory, research and practice. Newbury Park, CA: Sage Publications.

Wärneryd, B. et al. (1990): Att fråga: Om frågekonstruktion vid intervjuundersökningar och postenkäter. Stockholm: SCB.

Zimbardo, P.G. & Leippe, M.R. (1991): The Psychology of Attitude Change and Social Influence. Philadelphia: Temple University Press.

Øyen, E. (1990) (ed.): Comparative Methodology. London: Sage Publications.



## SFI-RAPPORTER SIDEN 2006

SFI-rapporter kan købes eller downloades gratis fra [www.sfi.dk](http://www.sfi.dk). Enkelte rapporter er kun udkommet som netpublikationer, hvilket vil fremgå af listen nedenfor.

- 06:01 Egelund, T.: *Sammenbrud i anbringelser*. 2006. 79 s. ISBN 87-7487-802-6. Kr. 65,00
- 06:02 Holt, H., Geerdsen, L.P., Christensen, G., Klitgaard, C. & Lind, M.L.: *Det kønsopdelte arbejdsmarked. En kvantitativ og kvalitativ belysning*. 2006. 250 s. ISBN 87-7487-804-2. Kr. 228,00.
- 06:03 Rosdahl, A.: *Kommunale aktiveringsprojekter med produktion*. 2006. 51 s. ISBN 87-7487-805-0. Kr. 50,00.
- 06:04 Christensen, E.: *Opvækst med særlig risiko. Indkredsning af børn med behov for en tidlig forebyggende indsats*. 2006. 92 s. ISBN 87-7487-806-9. Kr. 85,00.
- 06:05 Jørgensen, M.S., Holt, H., Hohnen, P. & Schimmel, G.: *Job på særlige vilkår. Overblik over viden på området*. 2006. 83 s. ISBN 87-7487-807-7. Kr. 75,00.
- 06:06 Rasmussen, M.: *Kontanthjælpsmodtageres gæld. Eftergivelse af offentlig gæld*. 2006. 68 s. ISBN 87-7487-808-5. Kr. 55,00.

- 06:07 Møller, S.S. & Rosdahl, A.: *Indvandrere i job. Marginalisering og beskæftigelse blandt ikke-vestlige indvandrere og efterkommere*. 2006. 171 s. ISBN 87-7487-809-3. Kr. 160,00.
- 06:08 Bengtsson, S. & Kristensen, L.K.: *Sårforborgens udlægning*. 2006. 96 s. ISBN 87-7487-810-7. Kr. 100,00.
- 06:09 Larsen, M.: *Fastholdelse og rekruttering af ældre. Arbejdspladsers indsats*. 2006. 101 s. ISBN 87-7487-813-1. Kr. 100,00.
- 06:10 Hestbæk, A.-D., Lindemann, A., Nielsen, V.L. & Christoffersen, M.N.: *Nye regler – ny praksis. Ændringerne i servicelovens børneregler 2001. Afslutningsrapport*. 2006. 265 s. ISBN 87-91247-80-2. Rapporten er udgivet af Styrelsen for Social Service. Kontakt denne eller send bestilling pr. e-mail til [bestilling@servicestyrelsen.dk](mailto:bestilling@servicestyrelsen.dk).
- 06:11 Olsen, H.: *Guide til gode spørgeskemaer. En manual*. 2006. 100 s. ISBN 87-7487-812-3. Kr. 100,00.
- 06:12 Bonke, J.: *Ludomani i Danmark. Udbredelsen af pengespil og problemspillere*. 2006. 79 s. ISBN 87-7487-811-5. Kr. 85,00.
- 06:13 Müller, M.M.: *Arbejds miljø og indvandrere. Erfaringer i forhold til rekruttering og fastholdelse*. 2006. 92 s. ISBN 87-7487-816-6. Kr. 90,00.
- 06:14 Hansen, H.: *Time Series of APW-Calculations - Module for Great Britain 1991-2004*. 2006. 83 s. ISBN 87-7487-815-8. Netpublikation.
- 06:15 Clausen, J., Heinesen, E. & Hussain, M.A.: *De nye kommuners rammevilkår for beskæftigelsesindsatsen*. 2006. 106 s. ISBN 87-7487-824-7. Netpublikation.
- 06:16 Christensen, G. & Christensen, S.: *Etniske minoriteter, frivilligt socialt arbejde og integration. Afdækning af muligheder og perspektiver*. 2006. 220 s. ISBN 87-7487-817-4. Netpublikation.
- 06:17 Schimmel, G.: *LO-dokumentation nr. 2/2006. Barrierer for kvinder i fagligt arbejde. En kvalitativ undersøgelse af årsagerne til kvinders lavere repræsentation i LO-fagbevægelsen*. 2006. 120 s. ISBN-10: 87-7735-770-1, ISBN-13: 978-87-7735-770-1. Kr. 20,00. Rapporten er udgivet af Landsorganisationen i Danmark.
- 06:18 Boje, T.P.: *Frivillighed og nonprofit i Danmark. Omfang, organisation, økonomi og beskæftigelse*. 2006. 275 s. ISBN 87-7487-821-2. Kr. 250,00.

- 06:19 Boje, T.P., Fridberg, T. & Ibsen, B. (red.): *Den frivillige sektor i Danmark. Omfang og betydning*. 2006. 172 s. ISBN 87-7487-822-0. Kr. 160,00.
- 06:20 Geerdsen, P.P. & Geerdsen, L.: *Fra aktivering til beskæftigelse. En gennemgang af aktiveringsindsatsen i det danske dagpengesystem*. 2006. 72 s. ISBN 87-7487-818-2. Kr. 70,00
- 06:21 Jespersen, C.: *Socialt udsatte børn i dagtilbud*. 2006. 108 s. ISBN 87-7487-835-2. Kr. 100,00
- 06:22 Christensen, V.T.: *Uørt? Betydningen af nedsat børselse for arbejdsmarkeds-tilknytning og arbejdsliv*. 2006. 254 s. ISBN 87-7487-823-9. Kr. 248,00.
- 06:23 Jensen, T.G., Schmidt, G., Jareno, K.N. & Roselius, M.: *Indsatser mod arerelateret vold*. 2006. 185 s. ISBN: 87-7487-825-5. Netpublikation.
- 06:24 Miiler, M.M., Høgelund, J. & Geerdsen, P.P.: *Handicap & beskæftigelse. Udviklingen mellem 2002 og 2005*. 2006. 128 s. ISBN 87-7487-826-3. Kr. 110,00.
- 06:25 Christensen, E. & Andersen, K.V.: *Livsvilkår for børn med familie på danske asylcentre*. 2006. 120 s. ISBN: 87-7487-827-1. Kr. 120,00.
- 06:26 Rostgaard, T.: *Oplysning om demens. En evaluering af Socialministeriets pulje til oplysning om demens*. 2007. 65 s. ISBN: 978-87-7487-829-2. Kr. 60,00.
- 06:27 Bengtsson, S. & Nemli, A.: *Oplevelsen af MST. Forældres, unges og terapeuters erfaringer med Multisystemisk Terapi*. 2006. 136 s. ISBN: 87-7487-830-1. Kr. 140,00.
- 06:28 Stigaard, M.V., Sørensen, M.F., Winter, S.C., Frisberg, N. & Henriksen, A.C.: *Kommunernes beskæftigelsesindsats*. 2006. 113 s. ISBN 87-7487-832-8. Kr. 90,00.
- 06:29 Madsen, M.B., Mortensøn, M.D. & Rosdahl, A.: *Arbejdsmarkedsparat eller ej? En kvalitativ undersøgelse af visitationen af kontanthjælpsmodtagere i ti kommuner*. 2006. 109 s. ISBN 87-7487-833-6. Kr. 100,00.
- 06:30 Rosdahl, A. & Petersen, K.N.: *Modtagere af kontanthjælp. En litteraturoversigt om kontanthjælpsmodtagere og den offentlige indsats for at hjælpe dem*. 2006. 87 s. ISBN 87-7487-834-4. Kr. 65,00.
- 06:31 Deding, M & Jakobsen, V.: *Indvandreres arbejdsliv og familieliv*. 2006. 101 s. ISBN 87-7487-836-0. Netpublikation.

- 06:32 Deding, M., Lausten, M. & Andersen, A.R.: *Børnefamiliers balance mellem familie- og arbejdsliv*. 2006. 139 s. ISBN 87-7487-837-9. Netpublikation.
- 06:33 Hansen, H.: *Time Series of APW-Calculations - Module for Denmark 1994-2005*. 2006. 121 s. ISBN 87-7487-838-7. Netpublikation.
- 06:34 Christensen, E.: *Uledsagede asylansøgerbørn*. 2006. 64 s. ISBN 87-7487-840-9. Kr. 65,00.
- 06:35 Christensen, V.T.: *Hard of Hearing? Hearing problems and working life*. 2006. 49 s. ISBN 87-7487-823-9. Engelsk sammenfatning af rapporten *Uhørt?* Netpublikation.
- 07:01 Damgaard, B. & Boll, J.: *Opfølgning på sygedagpenge – Del I. Kommuners, lægers, og virksomheders erfaringer med de nye regler*. 2007. 116 s. ISBN 978-87-7487-842-1. Kr. 100,00.
- 07:02 Bach, H.B. & Petersen, K.N.: *Kontanthjælpsmodtagerne i 2006. En surveyundersøgelse af matchkategorier, arbejde og økonomi*. 2007. 146 s. ISBN 978-87-7487-843-8. Kr. 110,00.
- 07:03 Sivertsen, M.: *Hvordan virker indsatsen mod negativ social arv? Gennemgang og analyse af 54 projektevalueringer*. 2007. 55 s. ISBN 978-87-7487-844-5. Kr. 60,00.
- 07:04 Jespersen, S.T., Junge, M., Munk, M.D. & Olsen, P.: *Brain drain eller brain gain? Vandringer af højtuddannede til og fra Danmark*. 2007. 64 s. ISBN 978-87-7487-846-9. Netpublikation.
- 07:05 Benjaminsen, L.: *Storbypuljen – Indsatser for socialt udsatte. Ideer og erfaringer*. 2007. 47 s. ISBN 978-87-7487-847-6. Kr. 60,00. Pjece.
- 07:06 Miiller, M.M., Havn, L., Holt, H. & Jensen, S.: *Virksomheders sociale engagement. Årbog 2006*. 2007. 178 s. ISBN 978-87-7487-848-3. Kr. 180,00.
- 07:07 Madsen, M. B., Filges, T., Hohnen, Jensen S. & Nærvig Petersen, K.: *Vil De gerne have et arbejde?* 2007. 194 s. ISBN 978-87-7487-849-0. Kr. 175,00.
- 07:08 Nielsen, C., Benjaminsen, L., Dinesen P. T. & Bonke, J.: *Effektmåling*. 2007. 180 s. ISBN 978-87-7487-850-6. Netpublikation.
- 07:09 Boesby, D.: *At oplyse om demens. Ideer og inspiration*. 2007. 18 s. Netpublikation

- 07:10 Graversen, B.K., Damgaard, B. & Rosdahl, A.: *Hurtigt i gang. Evaluering af et forsøg med en tidlig og intensiv beskæftigelsesindsats for forsikrede ledige*. 2007. 107 s. ISBN 978- 87-7487-851-3.
- 07:11 Thorsager, L., Børjesson, E., Christensen, I. & Pihl, V.: *Metoder i socialt arbejde. Begreber og problematikker*. 2007. 128 s. ISBN 978-87-7487-852-0. Kr. 120,00.
- 07:12 Hohnen, P., Mortensøn, M.D. & Klitgaard, C.: *Den korteste vej til arbejdsmarkedet. En kvalitativ undersøgelse af indsatsen over for ikke-arbejdsmarkedsparede ledige*. 2007. 145 s. ISBN: 978-87-7487-854-4. Kr. 138,00.
- 07:13 Rostgaard, T.: *Begreber om kvalitet i aldreplejen. Temaer, roller og relationer*. 2007. 225 s. ISBN 978-87-7487- 855-1. Kr. 218,00.
- 07:14 Bonke, J.: *Ludomani i Danmark. Faktorer af betydning for spilleproblemer*. 2007. 90 s. ISBN 978-87-7487-853-853-7. Kr. 90,00.
- 07:15 Andersen, D. & Højlund, O.: *Interview med 11-årige. Erfaringer fra et web-baseret pilotprojekt*. 2007. 121 s. ISBN: 978-87-7487-857-5. Kr. 110,00.
- 07:16 Dahl, K.M: *Udsatte børns fritid – et litteraturstudie*. 2007. 85 s. ISBN: 978- 87-7487-858-2. Netpublikation.
- 07:17 Vinther, H: *Rundt om forebyggelses- og sundhedscentre. Muligheder og barrierer for udvikling af forebyggelses- og sundhedscentre i Danmark*. 2007. 78 s. ISBN: 978-87-7487-860-5. Netpublikation
- 07:18 Fridberg, T. & Jæger, M.M.: *Frivillige i Hjemmeværnet*. 2007. 97 s. ISBN: 978- 87-7487-861-2. Kr. 90,00.
- 07:19 Høgelund, J. & Larsen, B.: *Handicap og beskæftigelse. Udviklingen mellem 2005 og 2006*. 2007. 39 s. ISBN: 978-87-7487-864-3. Netpublikation.
- 07:20 Larsen, B., Müller M.M. & Høgelund, J.: *Handicap og beskæftigelse. Regionale forskelle*. 2007. 65 s. ISBN: 978-87-7487-865-0. Netpublikation.
- 07:21 Jørgensen, M.: *Danskernes pensionsopsparinger. En deskriptiv analyse*. 2007. 238 s. ISBN: 978- 87-7487-866-7. Kr. 198,00.
- 07:22 Benjaminsen, L. & Christensen, I.: *Hjemløshed i Danmark 2007. National kortlægning*. 2007. 159 s. ISBN 978- 87-7487-867-4. Kr. 148,00.

- 07:23 Nielsen, V.L. & Ploug, N.: *Når politik bliver til virkelighed. Festskrift til professor Søren Winter*. 2007. 241 s. ISBN: 978- 87-7487-868-1. Kr. 229,00.
- 07:24 Egelund, T. & Vitus, K.: *Sammenbrud i anbringelser af unge. Risikofaktorer hos unge, forældre, anbringelsessteder og i sagsbehandlingen*. 2007. 71 s. ISBN: 978- 87-7487-869-8. Kr. 75,00.
- 07:25 Ploug, N.: *Socialt udsatte børn. Identifikation, viden og handlemuligheder i daginstitutioner*. 2007. 48 s. ISBN: 978- 87-7487-870-4. Kr. 50,00.
- 07:26 Olsen, H: *Konstruktion og kvalitetssikring af multisurveydata*. 2007. 175 s. ISBN: 978-87-7487-871-1. Netpublikation.



# KONSTRUKTION OG KVALITETSSIKRING AF MULTISURVEYDATA

Multisurveyundersøgelser, dvs. surveyundersøgelser på tværs af landegrænser og kulturer, har for alvor vundet indpas i løbet af de sidste 25 år. Denne rapport undersøger, hvordan data konstrueres og kvalitetssikres i multisurveys.

Rapporten, der er udarbejdet af adj. docent Henning Olsen, er baseret på et omfattende studie af metodelitteraturen om multisurveys, som han sammenholder med litteraturen om monosurveys. Studiet leder til en kritisk diskussion af forskellene mellem de to undersøgelsesmetoder og deres begrebsapparater. Forfatteren påviser, at flere centrale metodiske begreber – især inden for multisurveyforskningen – er vagt og modsætningsfyldt defineret, og han efterlyser derfor et mere dækkende og veldefineret begrebsapparat.