Jacobs University Bremen

Questionnaire Design in Attitude and Opinion Research: Current State of an Art

Petra Lietz

Priorisierung in der Medizin
FOR 655 Nr. 13 / 2008

JACOBS UNIVERSITY
Campus Ring 1
28759 Bremen
Germany
www.jacobs-university.de

Die Reihe „Priorisierung in der Medizin“ umfasst Arbeits- und Forschungsberichte der DFG Forschergruppe FOR655 „*Priorisierung in der Medizin: eine theoretische und empirische Analyse unter besonderer Berücksichtigung der Gesetzlichen Krankenversicherung (GKV).*

Die Berichte und weitere Informationen zu der Forschergruppe können abgerufen, werden unter:

http://www.for655.de oder http://www.priorisierung-in-der-medizin.de

The series „Priorisierung in der Medizin“ consists of working papers and research reports of the DFG (Deutsche Forschungsgemeinschaft, i.e., German Research Foundation) Research Group FOR655 „*Priorisierung in der Medizin: eine theoretische und empirische Analyse unter besonderer Berücksichtigung der Gesetzlichen Krankenversicherung (GKV).*(*Prioritizing in Medicine: A Theoretical and Empirical Analysis in Consideration of the Public Health Insurance System*)

Reports and further information can be found at

http://www.for655.de or http://www.priorisierung-in-der-medizin.de

Impressum:

JACOBS UNIVERSITY

Campus Ring 1
28759 Bremen
Germany
www.jacobs-university.de

ISSN 1866-0290

www.for655.de | www.priorisierung-in-der-medizin.de

Questionnaire Design in Attitude and Opinion Research: Current State of an Art

Petra Lietz*
Jacobs-University Bremen gGmbH

> The world is full of well-meaning people who believe that anyone who can write plain English and has a modicum of common sense can produce a good questionnaire. This book is not for them.
> Oppenheim (1966) - Preface to the first edition of Questionnaire design and attitude measurement.

1. Introduction

One of the components of the work schedule for the research group FOR655 "Prioritizing in medicine: A theoretical and empirical analysis in consideration of the public health insurance system" is to design and administer a questionnaire to a nationally representative probability sample of the German adult population in order to obtain information regarding the attitudes and opinions of Germans towards prioritizing services and treatments in medicine and to examine factors that influence those attitudes and opinions. In order to inform this process, the current paper is a review of research into various aspects of questionnaire design with particular emphasis on question wording and question order as well as on a number of issues concerning response scales, such as the number of response options, the labeling of response options and the desirability or otherwise of including a 'don't know' option.

Prof. Dr. Petra Lietz
School of Humanities and Social Sciences
Jacobs University Bremen gGmbH
Campus Ring 1
D-28759 Bremen
phone: 0421/200 3431
e-mail: a.diederich@jacobs-university.de

What this paper explicitly does not address are potential effects of different means of questionnaire administration on responses obtained from participants. In the project, the decision has been taken to administer the questionnaire in a computer assisted face-to-face interview (CAPI) situation as this is the highest standard of interview practice in survey research (ADM, 1999). Thus, research regarding the possible effects on the quality of responses and the quality of the obtained sample as a result of different means of questionnaire administration, such as face-to-face, mail, online or telephone administration or of interviewer characteristics such as age, gender or ethnicity are not covered in this paper.

Questions

The following interchange might serve to illustrate the importance of question wording and its relationship to the responses obtained:

"Two priests, a Dominican and a Jesuit, are discussing whether it is a sin to smoke and pray at the same time. After failing to reach a conclusion, each goes off to consult his respective superior. The next week, they meet again.
Dominican: „Well what did your superior say"?
Jesuit: „He said it was all right".
Dominican: „My superior says it was a sin".
Jesuit: „What did you ask him"?
Dominican: „I asked him if it was all right to smoke while praying".
Jesuit: „Oh! I asked my superior if it was all right to pray while smoking".
Sudman, S., & Bradburn, N.M. (1989, p.8)

The traditional survey model (Foddy, 1993) considers a survey as a complex communication process. First, agreement has to be reached as to what to ask within a framework or model encompassing the research questions and hypotheses to be addressed and tested by the information obtained from the study. Second, researchers or interviewers encode their request for information in a carefully standardized physical stimulus, the question at the beginning of the process. Respondents subsequently decode this stimulus and encode an answer which is usually expressed in terms of a standardized format which was previously encoded by the researcher. Finally, the researchers or interviewers decode this response and proceed to analyzing the information and drawing some form of conclusion from the analyses. This conceptualization of a survey as a communication process by Foddy (1993) and others (e.g. Cannell, Miller & Oksenberg, 1981; Tourangeau, Rips & Rasinski, 2000) not only focuses on its actors and their actions but also draws attention to all the points in the process where problems in or even break-downs of communication can occur. Hence, in this paper, consideration is given first to questions and the various possibilities of their encoding and decoding, followed by a discussion of issues involved in the encoding and decoding of responses.

Questions

Brace (2004) has emphasized the importance of question encoding to the success of the communication process, particularly in market research, which has to be able to successfully tune into the language of respondents that are diverse in terms of gender, age as well as level of education, occupation and income. Therefore, the research reported below is focused on best practice as regards question length, question wording and question order in order to avoid negative impact on sample quality due to non-response – which has been shown to increase over time (deLeeuw & deHeer 2002) - or on data accuracy due to respondents' misinterpretation of or deliberate lying in answer to questions. It should be kept in mind that good practice in terms of these issues is of particular importance in international research as it assists in reducing the impact of difference in culture and language on survey results (Brislin 1986; Smith 2003)

Question length

The general advice is to keep questions or statements as short as possible (Dillmann 2000; Fink 2003; Foddy 1993) with a maximum number of 16 (Brislin 1986) to 20 (Oppenheim 1992) words per sentence whereby questions can consist of more than one sentence.

In addition, Blair et al. (1977) and Andrews (1984) report increased data quality if questions or groups of questions concerning the same topic are preceded by a medium-length introduction (16 to 64 words, Andrews 1984; 30 words, Blair 1977). According to evidence reported by Oksenberg and Cannell (1977, p. 342) and Jabine (1989), somewhat longer questions lead to more accurate reporting as they may convey the idea that the task is important and deserves serious effort.

Grammar

Brislin (1986) as well as Dillman (2000) argue to keep the grammatical complexities to a minimum. Thus, questions should employ the active rather than the passive voice, repeat nouns instead of using pronouns and avoid possessive forms in order minimize the cognitive demands on respondents order to free up mental capacity to think about their response.

Specificity and simplicity

Another means of reducing the cognitive load on respondents stems from using specific rather than general terms (Brislin 1986; Dillmann 2000), breaking down more complex questions into simpler ones (Jobe & Mingay 1989) and avoiding words that indicate vagueness, such as "probably", "maybe", or "perhaps" (Brislin 1986; Dillmann 2000). Belson (1981) and Foddy (1993) also advise against the use of hypothetical questions concerning respondents' future behaviours. Instead, they recommend the use of vignettes

or alternative scenarios when seeking reactions to issues that are outside the realm of the past or present.

Many studies (Oksenberg & Cannell 1977; Rockwood et al. 1997; Tourangeau et al. 2000) as well as the meta-analysis of Sudman and Bradburn (1974) show that the invalidity of responses due to cognitive overload increases where recall of events is involved that have occurred not in the immediate past (i.e. more than a week ago) whereby the invalidity of responses depends on the importance of the event (e.g. visit to the GP vs. hospitalization; minor vs. major house repairs).

Social desirability (SD)

The merit of simplicity in question wording is emphasized by Foddy (1993) who labels the undesired off-putting effect of poorly worded questions on respondents "question threat". He adds that the use of difficult vocabulary either in questions or instructions leads to respondents feeling stupid or uneducated and increases the probability of obtaining "don't know" or socially desirable responses. Socially desirable responses can lead to answers that inaccurately reflect respondents' actual behaviours in a number of ways. First, respondents might choose to select a certain position that is thought to be one that is favoured by society (e.g. not to smoke or drink, to do exercise). As a consequence, particularly in medical research, people tend to underreport unhealthy lifestyle practices and over-report healthy ones (Brace 2004). Second, because of the social prestige which is attached to the act of uttering an opinion and the corresponding negative evaluation associated with the lack thereof (Leverkus-Brüning 1966) respondents think that they should be informed about certain issues (e.g. the EU constitution, climate change) and give responses conveying this impression instead of admitting ignorance. Third, Foddy (1993) states fear of being identified or revealing details about the private sphere or facts that are considered embarrassing, such as medical diagnoses of mental or genital diseases (Oksenberg & Cannell 1977) as reasons for respondents' giving socially desirable responses. It is mainly the first two aspects that are subsumed in Holtgraves' (2004 p. 161) definition of social desirability which "refers to a tendency to respond in self-report items in a manner that makes the respondent look good rather than to respond in an accurate and truthful manner".

In order to reduce respondents' propensity to give socially desirable answers especially on sensitive issues such as adultery, crime or drug use, Brace (2004), suggests indirect questioning, such as "What do you believe other people think about…" whereby the assumption is that respondents will more easily admit to views or behaviours that they think are not shared by the larger society by projecting their own views onto others. Or, if the issue involves knowledge that the respondent might not have, a phrasing such as "Have you had time yet to familiarize yourself with the new (EU) Constitution?" might facilitate the respondent's acknowledgement of his/her ignorance in this matter.

Another means of reducing respondents' propensity to give socially desirable answers is the use of the introductory phrase "Do you happen to know..." as Brace (2004) argues that this phrase allows respondents to think a bit longer in order to retrieve any knowledge they might have regarding the topic. Another beneficial aspect of this phrase is put forward by Bradburn, Sudman and Wansink (2004) who suggest this question wording in order to signal to participants with less firm attitudes or information bases that it is acceptable to volunteer a "don't know" response.

Other suggestions to reduce social desirability frequently include that questions (a) are worded as neutrally as possible, (b) propose values on a certain topic not only in one but different directions and (c) suggest the normalcy of socially deviant behaviour (Bortz & Döring 2003; Brace 2004; Oppenheim 1992; Scholl 2003). Diekmann (2003), however, has reported limited effects of such measures.

In addition, a number of instruments (e.g. the Edwards Social Desirability Scale Edwards 1957; the Balanced Inventory of Desirable Responding (BIDR) Paulhus 1984; Marlowe-Crowne Social Desirability Scale (MCDS), Crowne and Marlowe 1960) has been developed to measure SD in order to control for this tendency in subsequent analyses (Diekmann 2003; Seitz 1977). However, as much research has reported questionable validity and reliability (Leite & Beretvas 2005; Moorman & Podsakoff; 1992; Paulhus & Van Selst 1990; Paulhus 1991) for these instruments it seems that, although not much research has been done to test empirically the differences in SD that are likely to exist between countries (Stocké & Hunkler 2007), the proposed question wordings that are aimed at reducing the tendency to give socially desirable answers are preferable to the use of measures of social desirability that have questionable psychometric properties and would increase considerably the length of a questionnaire.

Double-barrelled questions

A number of authors recommend to avoid the ambiguity of so-called "double-barrelled" questions or statements that contain two different verbs or two different concepts. More specifically, Brislin (1986) mentions the use of two verbs in one question as being detrimental to the obtaining of accurate responses while Brace (2004), Fink (2003), Fowler (1992) and van der Zouwen (2000) extend it to the use of two concepts in one question. For example, the question "Do you have time to read the newspaper every day?" contains two aspects, namely 'having the time' and 'reading the paper every day', which is why the question "Do you read the newspaper every day?" followed by a question about reasons if this is (not) the case will be clearer. This question also illustrates that questionnaire designers have to be clear what it is that they want to obtain information on. At the start, the questionnaire designer might not have realized that the question contained two aspects, namely the behaviour and the reason for the behaviour. On a somewhat different aspect of double-barrelledness, a question such as "Should older people who smoke pay some of the costs related to a potential lung-cancer treatment themselves?" leaves open who the reference group is: Older people who do not smoke? Younger people who smoke? Younger people who do not smoke?

Negatively worded questions

The general advice is against the inclusion of negatively worded questions or statements (Belson 1981; Foddy 1993) as they have been found to take longer to process (Wason 1959; Weems et al. 2002) and have a greater likelihood of respondents to make mistakes (Dudycha & Carpenter 1973; Eifermann 1961), hence introducing an artificial methods effect into the response behaviour (DiStefano & Motl 2006). Foddy (1993) argues that this is particularly the case when the word "no/not" is used together with words that have a negative meaning. Thus, he suggests that the question "What is your view about the statement that conservationists should not be so uncooperative with the government" should be rephrased into "What is your view about the statement that conservationists should cooperate with the government" so that respondents do not have to go through a tiring process in order to deduce the meaning of the question. In addition, he emphasizes how quickly a question can turn into a double negative when taken together with the answer options, as is the case when respondents are asked to agree or disagree with the statement "Teachers should not be required to supervise students in the halls." O'Muircheartaigh et al. (2000, p. 22) confirmed the undesirability of negatively worded items as their analyses showed these to be less reliable than positively worded items. This evidence supports the notion that the introduction of negatively worded items into an item battery in order to balance it "introduces greater random error" although there is some evidence that this may not be the case of sophisticated item response techniques are used in the development of the scale (Bergstrom & Lunz 1998). An interesting aside in this context is the finding that more people are willing to respond "no" to "allowing" something (e.g. x-rated movies, cigarette advertisements) than to respond "yes" to "forbidding" it (Schumann & Presser 1977, 1978; Hippler & Schwarz 1986).

Adverbs of frequency

Another recommendation for clear question wording concerns the use of adverbs that indicate frequency. In an early study, Simpson (1944) asked people for 20 frequency adverbs to indicate the percentage of time this word meant that something occurred. He found the largest agreement for the terms "never" (0-2% of the time), "almost never" (3-5% of the time), "about as often as not" (48-50% of the time) "always" (98-100% of the time) and the largest difference in interpretation for the terms "frequently" (40-80% of the time) and "rather often" (45-80% of the time). Moreover, he found no frequency terms that were interpreted by people to indicate occurrences of between 20 and 50 per cent of the time. Similarly, Liechtenstein and Newman (1967) reported the smallest range in interpretation for the "middle-of-the–frequency-road" term "tossup" (45-52%) and the largest range in interpretation for the terms "predictable", "probable" and "possible" (all from 1-99%).

Since then, a general consensus has emerged that "frequently", "usually", and "regularly" have quite different meanings for different respondents and depending on the question content (Bradburn & Miles 1979; in Krumpal et al. 2008) as well as on the numeric values assigned if these terms are used as labels of a response scale (Schwarz,

Grayson & Knäuper 1998). To highlight this, Foddy (1993, p. 43) reported 445 interpretations of the word "usually" as the meaning assigned to the word varied depending on, for example, the type of activity or who was asked about the activity.

One solution to this problem is to offer participants more specific quantifiers in the response options. Therefore, "never or almost never", "once or twice a month", "once or twice a week" and "always or almost always" are used as response options to many of the questions asked in background questionnaires addressed at teachers and principals as part of internationally comparative studies in education (e.g. Mullis et al. 2007). In addition, as suggested by Bradburn and Sudman (1979), questions aimed at obtaining information regarding frequency of behaviour should include numeric reference points for a specified time period. Thus, a question about watching television should be worded "How many hours do you watch TV on a week-day (excluding week-ends)?" with response options such as "< 0.5 hours", "0.5 hours to < 1 hour", "1 hour to < 1.5 hours", "1.5 hours to < 2 hours", "2 hours to < 2.5 hours", ">2.5 hours". Of course, this requires accurate knowledge about the question topic to enable the appropriate specification of the time period in the question (Dillman 2000; Fink 2003) and the response categories offered as answers (Gaskell et al. 1994; Schwartz et al. 1985).

Question order

Question order effects arise when answering behaviour changes depending on the position of a question during the interview (Schumann & Presser 1996). They are problematic in that they not only threaten the validity of the results but also the generalizability of results to the population about which conclusions are sought to be drawn. Types of question order effects include effects of part-whole combinations, part-part combinations and salience.

Question order effects of part-whole combinations occur where one question is more general with respect to a certain concept while the other is more specific. Examples are questions about respondents' state of happiness in general and their happiness in marriage or respondents' views on abortion in general and on abortion for specific reasons. Systematic research into this issue has been inconclusive as regards the answering behaviour in response to specific questions. For the general question, however, results tend to show that the general question is more appropriately placed before the specific question. This is argued to be due to the fact that the specific question takes a certain aspect out of the concept (e.g. marital happiness from general happiness or severe disability for the concept of abortion) which, then, is removed in the respondents' mind if the general question is asked after the specific question (Schumann & Presser 1996).

Question order effects of part-part combinations arise where questions are asked at the same level of specificity and respondents adapt their answers as a result of normative consistency. Thus, in two questions on (a) whether or not US American reporters should be allowed into what was then the Soviet Union and (b) whether or not reporters from

the Soviet Union should be allowed to enter the USA, Schumann and Presser (1996) found agreement with the second question to be significantly greater if (a) preceded (b) than if (b) was asked before (a). The authors reported similar results for questions regarding allowing US citizens to join the British, French or German armies and vice versa in that agreement to allow foreigners into the US army was far higher if this question was asked second. Counter-evidence, however, emerged for experiments regarding the extent to which people thought lawyers or doctors served the public good as well as a question where respondents were asked for their self-placement into a social class before and after questions regarding their education and occupation. In neither case did a question order effect emerge. Thus, it appears that it depends on the topic as to whether or not question order effects arise for part-part combinations.

Question order effects as a result of salience are said to occur when response behaviour changes as a result of a topic having been raised as part of the questioning process, hence conveying importance of that topic to respondents (Schumann & Presser 1996). Gaskell et al. (1994) found that between 9 and 13 per cent more respondents reported annoyance with adverts and feeling unsafe if previous questions in the survey had touched on these topics.

Demographic questions about respondents such as age, education, income and marital status should come at the end of the questionnaire rather than at the beginning in order to avoid negative feelings about the provision of personal information impacting on the answering behaviour or participation (Converse & Presser 1986; Oppenheim 1992).

Responses

The second main area for discussion in the survey communication framework revolves around the responses that are given to answer questions. Here, the relevant issues pertain to the standardized format or response stimuli in the form of response categories or scales generated on the part of the researcher as well as the process of encoding on the part of the respondent.

Don’t know option

Probably the first central issue that needs to be addressed on the part of the researcher is whether all respondents should answer all questions or whether those respondents with little or no knowledge should be filtered out and not be asked certain question. A related issue is - in the context of a standardized interview that is conducted in person - either face-to-face or by telephone - whether response scales should offer a “don’t know” (DK) option either explicitly or record it only when it is volunteered. To investigate this issue, Schumann and Presser (1996) conducted 19 experiments that compared responses to questions on US foreign affairs, courts, governments, and leadership with and without an explicitly offered DK option. They found that the percentage of respondents choosing DK increased by between 22 and 25 percent which was in line with findings reported by Trometer (1996). This difference in percentages held regardless of the familiarity of

respondents with the question topic as, for example, a question regarding the Portuguese government with which respondents were less familiar increased by from 63.2 per cent to 87.9 per cent whereas the DK proportion in response to a question regarding the US American government increased by about the same amount from 15.2 to 37.6 per cent. Looking at it in a different way, about one fifth of respondents shifted from the DK option to a substantive response option (i.e. "agree" or "disagree") if the DK option was not explicitly offered.

To examine whether or not the explicit offering of a DK option altered the distributions for the substantive response categories, Schumann and Presser (1996) compared the proportions of respondents choosing the agree and disagree options after omitting the respondents who chose the DK option in the two response types. Results indicated a large significant difference regarding respondents' choice of substantive response options for only one of the 19 experiments.

Opinion floating

Schumann and Presser (1996, p. 118) label people who give a substantive response when the DK is not offered but who choose this option when it is offered floaters as these people seem to vary their responses depending on the response options on offer. To investigate the extent to which these may systematically differ from other respondents, the authors conducted further experiments. Their results showed that while, in general, less educated respondents tended to give more DK responses than more educated respondents, it was the latter group for which a higher percentage of DK was recorded when the question topic had virtually not been covered in the media. The authors argued that this showed that, for topics that were generally less widely known, more educated respondents were willing to admit ignorance whereas less educated respondents used information given by the question to develop a substantive response. The authors (1996, p. 160) concluded "whether filtered or standard questions should be used in a questionnaire would seem to depend on whether an investigator is interested mainly in an "informed opinion" on an issue or mainly in underlying disposition".

Opinion filtering

A more explicit way of filtering out respondents is to ask questions such as "Do you have an opinion on this or not?" or "Have you been interested enough to favour one side over the other?" While such questions are advocated by some as a means of excluding anyone who is "ignorant" on a particular issue, two things have to be kept in mind. First, respondent's self-identification as being "ignorant' might vary systematically as a consequence of question topic as well as respondents' characteristics such as gender and age. Second, a serious consequence of filtering out respondents is the impact on the representativeness of the sample, in particular where stronger filter questions are used (e.g. "Have you already thought sufficiently about XYZ so that you could form an opinion" instead of "Do you have an opinion on XYZ?") that lead to the overestimation of people without an opinion (Hippler Schwarz & Sudman 1987). A commonly used

rule-of-thumb in survey research (Martin Mullis Kennedy 2007) is to consider a sample as being not representative of the intended target population if information is obtained from less than 80% of originally selected participants.

Bishop et al. (1979) tested the hypothesis that filtering out respondents through specific questions did not make a difference to the magnitude of the correlations between attitude items. To this end, they examined responses to five studies of US American adults with comparable sample compositions in terms of age, gender, race and education. Correlational analyses between respondents' attitudes towards government responsibilities, legalization of marijuana and their self-reported location on the liberal-conservative continuum showed higher correlations when filtering questions were applied. The authors argued that this evidence supported their "non-attitude hypothesis" according to which higher correlations should emerge between political attitude items with a prior filter than for items without a prior filter since the former would exclude respondents without firm attitudes.

Evidence that runs counter to the hypothesis that less well-informed people have no attitudes on certain issues stems from such people being consistent in their response behaviour over time. Moreover, for the group of people with non-attitudes it could be anticipated that half of them would favour an issue and the other half would oppose an issue. However, in an experiment involving questions that asked about issues to which the general public was known to have had little, if any, exposure Schumann and Presser (1996) found that this was not the case. This substantiated the earlier assumption by Allport (1935, as cited in Schumann & Presser 1996) that people used their general attitudes to guide them in the evaluation of questions with unfamiliar content. The experiments also provided supportive evidence for this assumption in that substantive responses to less well-known issues were related in a systematic way to other items that asked about similar issues but whose content was more widely known. This combined evidence led the authors to conclude "the evidence […] narrows, if indeed it does not eliminate, the conceptual distinction between attitudes and non-attitudes" (Schumann & Presser 1996).

Number of response scale options

A number of authors (Brace 2004; Dillman 2000; Fink 2003; Mayer 2002) report that between five-point and seven-point scale response options are the most commonly used. The seven-point scale has been shown to be more reliable (Cronbach 1951) as it allows for greater differentiation of responses than the five-point scale (Alwin 1992; Finn 1972; Masters 1974) while not artificially increasing differentiation (Cox 1980; Porst 2000; Schwarz & Hippler 1991), as might be the case where more scale points are offered.

Other authors also report evidence that supports the use of longer response scales. Rodgers et al. (1992) who investigated the effect of scale length from two to ten response options found that the expected value of the validity coefficient increased by about 0.04 for each additional response option while Matell & Jacoby (1971) found no

such linear increase when comparing concurrent validity coefficients for scales with 2 to 19 response options. Alwin (1997) conducted a confirmatory factor analysis of concepts being measured on seven-point scales (labelled "satisfied" "dissatisfied" and "delighted" to "terrible") compared to concepts being measured by a number of 11-point "feeling thermometers". Results indicated that 11-point scales had consistently higher reliability and validity coefficients and lower invalidity coefficients.

Instead of relating the optimal length of response scales to the distribution of responses Foddy (1993) relates it to the content of the question. Thus, Foddy argues that shorter scales such five-point scales are preferable in situations where respondents are asked for absolute judgments. In contrast, he considers longer scales such as seven- to nine-point scales to be more appropriate in situations where more abstract judgments are sought from respondents.

Odd or even number of response scale options

In addition to the question regarding the optimal number of response scale options, a decision has to be made whether to offer respondents an even or an odd number of response scale options. This implies a decision on whether or not to offer a - usually neutral - "middle" option that allows respondents not to commit themselves to a direction in their opinion or attitude.

Much research (Garland 1991; O'Muircheartaigh 2000; Kalton et al. 1980; Krosnick & Helic 2000; Schumann and Presser 1996) has shown that a middle alternative attracts between six and 23 per cent of respondents when it is offered, although, contrary to popular belief, the tendency to choose a middle option is not generally depending on age, education or gender (Kalton et al. 1980). O'Muircheartaigh et al. (2000) proceeded to examine in detail the shift in response distribution that occurred as a result of the inclusion or omission of the middle alternative. They found that the omission of the middle alternative increased responses to the DK option only slightly by one to two per cent. In addition, results showed a slightly higher increase for the weak agree/disagree responses (8.5%) than for the more extreme agree/disagree responses (4.1%) if the middle option was omitted. This latter result was also in line with the results of an experiment by Schumann and Presser (1996) who found that the introduction of moderate alternatives in a question about liberal-conservatism (i.e. "somewhat liberal" and "somewhat conservative") attracted more respondents from the middle alternative than from the extreme response alternatives.

O'Muircheartaigh et al. (2000) also examined the "satisficing" hypothesis initially put forward by Krosnick (1991). Krosnick (1991) hypothesizes that because many survey participants are likely to have low motivation and may find the task of responding difficult and exhausting they select the response alternative which involves the least amount of thinking and justifying. One of the implications of the satisficing hypothesis is the expectation that an omission of the middle alternative results in people "reporting meaningful attitudes that they would otherwise not have bothered to describe"

(O'Muircheartaigh et al. 2000, p. 20). Results of O'Muircheartaigh et al.'s (2000) analysis, however, did not support this hypothesis. Instead, response scales without the middle point had lower validity and higher random error variance, indicating that people randomly chose other available response options when the middle option was not available.

O'Muircheartaigh et al.'s (2000) analyses also revealed some insights into a phenomenon called "acquiescence" (Lenski & Leggett 1960) which refers to the tendency of respondents to agree with any statement regarding of its content, is the result of satisficing. Their analyses confirmed other evidence that such an effect exists (Smith 2004) and highlighted that a two-factor model consisting of (a) the actual attitude towards science and technology and (b) acquiescence was the model that fitted the data best.

Labelling of response scale options

Decisions regarding the labelling of response scale options include whether to use numbered scales that are unipolar (e.g. "On a scale from 0 to 10…") or bipolar (e.g. "Consider a scale from -5 to +5…") or verbal scales (e.g. "agree, slightly agree, neither agree nor disagree, slightly disagree, disagree" or "Would you say that you're very happy, pretty happy or not too happy these days?") and whether to label all responses options or only some of the response scale options.

Evidence from a number of studies (Fowler 1995; O'Muircheartaigh Gaskell and Wright 1995; Schwarz Knauper Hippler Noell-Neumann Clark 1991) have shown a greater likelihood for respondents to choose positive ratings on the bipolar scale than ratings of greater than five on the unipolar response scale. This finding held for topics as different as the entertainment value of movies and TV to general life satisfaction.

O'Muircheartaigh et al. (1995) further investigated the effect of differential response scale labelling not only in terms of numbers but also verbal anchors. They reported that the explicit mentioning of the verbal anchors made a difference to responses only to the "0" to "10" scale in that the "0" response option was chosen whereas it was not selected when the verbal anchors were not explicitly mentioned.

In a second experiment, O'Muircheartaigh et al. (1995) compared four combinations of unipolar and bipolar numerical and verbal scales. First, they found that midpoint of both numerical scales (i.e. -5 to +5 and 0-10) was chosen far more frequently (by about 30% of respondents) for the bipolar verbal anchors (i.e. the advertising authority should be given "much less power" and "given much more power") than the unipolar verbal anchors (i.e. not given any more power" and "given much more power" chosen by about 20% of respondents). Second, the lowest scale points ("0" and "-5" respectively) were chosen far more frequently if the verbal anchors were unipolar (16% and 15% respectively) than when they were bipolar (7% and 6% respectively).

A number of studies have investigated the verbal labelling of response scales tapping into the "good-bad" continuum (Mosier 1941; Myers & Warner 1968; Vidali 1975; Wildt & Mazis 1978). Results indicated that the words "disgusting", "unsatisfactory", "neutral", "desirable" and "excellent" produced normal distributions that overlapped little whereas words such as "acceptable", "important" and "indifferent" polarized respondents. In addition, participants with very different backgrounds rated "fantastic" and excellent (Mittelstaedt 1971; Myers & Warner 1968) to be the most positive adjectives and "horrible" and "terrible" to be the most negative adjectives. Finally, for the term "delightful", respondents varied the least whereas the term "unacceptable" respondents varied the most.

Other research has investigated the effects of so-called "multiplying adverbs" or "intensifiers" (e.g. "slightly", "rather" "extremely") on response distributions. Thus, Cliff (1959) asked respondents to rate the favourableness or otherwise of adjectives (e.g. "respectable", "mediocre") with and without such adverbs. He found that "slightly" and "somewhat" had the smallest intensifying effect, "very" and "extremely" had the largest intensifying effect while "pretty" and "quite" were closest to the meaning of an adjective without an intensifier. Similarly, Worcester and Burns (1975) found that adding "slightly" to the two moderate points of a five-point agree-disagree scale decreased the overlap of answers. O'Muircheartaigh et al. (1993) examined the effect of adding (a) "really" to a question on the frequency of feeling annoyed by an advert on television, (b) "very" to a question regarding the frequency of feeling unsafe around the neighbourhood in which they live, and (c) "extreme" to a question on the frequency of experiencing physical pain. While the effect on the distribution of responses for (a) and (b) was negligible a significant shift in distribution occurred when "extreme" was added to the question regarding physical pain. Indeed, only 38 per cent of respondents were aware that an intensifier had been used in the question about television advertisements whereas 75 per cent of respondents were aware of the use of the word "extreme" in the question regarding physical pain. This could, however, be a result of respondents assigning a much higher intensity to "extremely" than "very" as was demonstrated by Bartram and Yelding (1973).

Order of response options

Foddy (1993) has outlined a number of response options effects, including the primacy and recency effect as well as the effects of shifting frames of reference. The primacy effect refers to the assumption that respondents will select earlier alternatives more frequently than later alternatives, especially when alternatives are presented on "show cards". The recency effect is said to apply when respondents select the later alternatives and is thought to apply mainly when respondents only hear the alternatives. The phenomenon of shifting frames of reference refers to the possibility that the selection of an certain alternative depends on whether the "more favourable" alternatives are presented earlier or later. Schumann and Presser (1996) examined these effects in detail and found some evidence of a recency effect but only for unusual topics and long-

winded questions as well as of a primacy effect for very long lists that include 16 alternatives.

Finally, while Fink (1995) asserts that the direction of the response options is negligible in most situations, Bradburn et al. (2004) recommend to put those options first (i.e. on the left) that convey less socially desirable responses to prevent respondents from making a choice without having read all available options.

Conclusion

From the above review of research into questionnaire design, a number of recommendations emerge:

Questions should be constructed to be as clear, simple, specific and relevant for the study's research aims as possible;
Questions should focus on current attitudes and very recent behaviour;
More general questions should precede more specific questions;
Vague quantifiers such as "frequently", "usually", and "regularly" should be avoided. Instead, carefully pre-tested response options should specify the number of times per appropriate period (e.g. day, week, month, year) of an event or behaviour;
A desirable Likert-type response scale length ranges from 5 to 8 response options;
The inclusion of a middle option increases the validity and reliability of a response scale slightly;
The numerical scale should be unipolar with matching verbal labels as anchors at both ends of the scale;
"Extremely" and "not at all" can serve as most effective verbal intensifiers;
All numeric labels should be shown to respondents;
Numeric and verbal anchors (=endpoints) should be mentioned explicitly;
A "don't know" option should be recorded if volunteered whereby interview instructions should be such that interviewers are not to encourage respondents to choose a substantive response options if they hesitate;
Demographics questions should be put at the end of the questionnaire.

Of course, adherence to these recommendations for questionnaire design will only serve to go some way in the development of a questionnaire that is of high quality. The next step in the questionnaire design process will be the cognitive (e.g. Jobe & Mingay 1998; Willis 2005) and quantitative piloting (e.g. Litwin 2003; DeVellis 2003; Presser & Blair 1994) of the questionnaire in order to allow for an evaluation in terms of its acceptance and understanding by members of the intended target population and an analysis of the psychometric properties (e.g. Andrich 1978; von der Linden & Hambleton 1997; Nunnally & Bernstein 1994; Wright & Masters 1982) of its constituent questions and scales.

*Author's note

The assistance of Ognyan Seizov in the retrieval and summarizing of the literature for this paper is gratefully acknowledged.

References

ADM - Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (October, 1999). Standards for quality assurance in market and social research. Retrieved http://www.adm-ev.de/pdf/QUALI.PDF on 04/06/08.

Allport, G. (1935). Attitudes. In C.M. Murchison (ed.) Handbook of social psychology (pp. 798-844). Worchester, MA: Clark University Press. Cited in Schumann & Presser 1996, op. cit.

Alwin, D.F. (1992) Information transmission in the survey interview: Number of response categories and the reliability of attitude measurement. Sociological Methodology, 22, pp. 83-118.

Alwin, D.F. (1997) Feeling thermometers vs. 7-point scales: Which are better?. Sociological Methods and Research, 25, pp. 318-40.

Andrews, F. (1984) Construct validity and error components of survey measures: A structural modeling approach. Public Opinion Quarterly, 48, 2, pp. 409-442.

Andrich, D. (1978) Scaling attitude items constructed and scored in the Likert tradition. Educational and Psychological Measurement, 38, pp. 665-80.

Bartram, P. & Yelding, D. (1973) The development of an empirical method of selecting phrases used in verbal rating scales: A report on a recent experiment. Journal of The Market Research Society, 15, pp. 151-156.

Belson, W. A. (1981). The design and understanding of survey questions. Aldershot, England: Gower.

Bergstrom, B.A. & Lunz, M.E. (1998) Rating scale analysis: Gauging the impact of positively and negatively worded items. Paper presented at the Annual Meeting of the American Educational Research Association, April 13-17.

Bishop, G.F., Oldendick, R.W., Tuchfarber, A.J. & Bennett, S.E. (1979) Effects of opinion filtering and opinion floating: Evidence from a secondary analysis. Political Methodology, pp. 293-309.

Blair, E., Sudman, S., Bradburn, N. & Stocking, C. (1977) How to ask questions about drinking and sex: Response effects in measuring consumer behavior. Journal of Marketing Research, 14, pp. 316-21.

Bortz, J. & Döring, N. (2003) Forschungsmethoden und Evaluation für Sozialwissenschaftler. Berlin, Heidelberg: Springer.

Brace, I. (2004) Questionnaire design. How to plan, structure and write survey material for effective market research. London: Kogan Page.

Bradburn, N. & Miles, C. (1979) Vague quantifiers. Public Opinion Quarterly, 43, pp. 92-101. In Krumpal et al. (2008). Op.cit.

Bradburn, N. & Sudman, S. (1979) Improving interview method and questionnaire design. San Francisco: Jossey-Bass.

Bradburn, N., Sudman, S. & Wansink, B. (2004) Asking questions. The definitive guide to questionnaire design—For market research, political polls, and social and health questionnaires. San Francisco, CA: Jossey-Bass.

Bradburn, N. (1983) Response effects. In Rossi, P., Wright, J. & Anderson, A. (Eds.), Handbook of survey research (pp. 289-328) New York: Academic Press.

Brislin, R.W. (1986). The wording and translation of research instruments. In W.J. Lonner & J.W. Berry (eds.), Field methods in cross-cultural research (pp. 137-164). Newbury Park, CA: Sage.

Cannell, C.F., Miller, P.V. & Oksenberg L. (1981) Research on interviewing techniques. In: S. Leinhardt (ed.) Sociological methodology. San Francisco, Cal.: Jossey Bass.

Cliff, N. (1959) Adverbs as multipliers. Psychological Review, 66, pp. 27-44.

Converse, J. & Presser, S. (1986) Survey questions. Handcrafting the standard questionnaire. London: Sage.

Cox III, E.P. (1980) The optimal number of response alternatives for a scale: A review. Journal of Marketing Research, 17, pp. 407-422.

Cronbach, L.J. (1951) Coefficient alpha and the internal structure of tests. Psychometrica, 16, 93-96.

Crowne, D. & Marlowe, D. (1960) A new scale of social desirability independent of psychopathology. Journal of Consulting Psychology, 24, pp. 963-968.

deLeeuw, E. & deHeer, W. (2002) Trends in household survey nonresponse: A longitudinal and international comparison. In R. Groves, D. Dillman, J. Eltinge & R. Little (eds.) Survey non-response (Chapter 3, pp. 41-54) New York: Wiley.

DeVellis, R. F. (2003) (2nd ed.) Scale development. Theory and application. Thousand Oaks, Cal.: Sage.

Diekmann, A. (2003) Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbeck bei Hamburg: Rowohlt.

DiStefano, C. & Motl, R.W. (2006) Further investigating method effects associated with negatively worded items on self-report surveys. Structural Equation Modeling: A Multidisciplinary Journal,13, 3, pp. 440 — 464

Dillman, D. (2000) Mail and Internet surveys. The tailored design method. New York: John Wiley & Sons, Inc.

Dudycha, A.L. & Carpenter, J.B. (1973) Effects of item format on item discrimination and difficulty. Journal of Applied Psychology, 58, pp. 116-121.

Edwards, A. (1957) The social desirability variable in personality assessment and research. New York: Dryden Press.

Eifermann, R.R. (1961) Negation: A linguistic variable. Acta Psychologica, 18, pp. 258-273.

Fink, A. (2003) How to ask survey questions. Thousand Oaks, Cal. London: Sage.

Finn, R. (1972). Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. Educational and Psychological Measurement, 32, 2, 255-65.

Foddy, W. (1993) Constructing questions for interviews and questionnaires. Theory and practice in social research. Cambridge, UK: Cambridge University Press.

Fowler, F. (1995) Improving survey questions. Design and evaluation. London: Sage.

Fowler, F. (1992) How unclear terms affect survey data. Public Opinion Quarterly, 56, 2, pp. 218-31.

Garland, R. (1991) The mid-point on a rating scale: Is it desirable? Marketing Bulletin, 2, 66-70.

Gaskell, G.D., O'Muircheartaigh, C.A. & Wright, D.B. (1994) Survey questions about the frequency of vaguely defined events: The effects of response alternative. Public Opinion Quarterly, 58, 2, pp. 241-254.

Hippler, H.-J. & Schwarz, N. (1986) Not forbidding isn't allowing: The cognitive basis of the forbid-allow asymmetry. The Public Opinion Quarterly, 50, 1, pp. 87-96.

Hippler, H.-J., Schwarz, N. & Sudman, S. (1987) (eds.) Social information processing and survey methodology. New York: Springer.

Holtgraves, T. (2004) Social desirability and self-reports: Testing models of socially desirable responding'. PSPB, 30, 2, pp. 161-172.

Hunt, J.H., Domzal, T.J. & Kernan, J.B. (1982) Causal attributions and persuasion: The case of disconfirmed expectancies. In: A.A. Mitchell (ed.), Advances in consumer research (pp. 287–292). Ann Arbor, MI: Association for Consumer Research.

Jabine, T.B. (1987) Reporting chronic conditions in the National health Interview Survey: A review of tendencies from evaluation studies and methodological test. Vital and Health Statistics, Series 2, 105. Washington, D.C.: Government Printing Office.

Jabine, T.,B. Straf, M.L., Tanur, J.M. & Tourangeau, R. (eds.). (1984) Cognitive aspects of survey methodology: Building a bridge between disciplines. Washington, D.C.: National Academy Press.

Jobe, J. & Mingay, D. (1989) Cognitive research improves questionnaires. American Journal of Public Health, 79, 8, pp. 1053-1055.

Kalton, G., Robert, J. & Holt, D. (1980) The effects of offering a middle response option with opinion questions. The Statistician, 29, pp. 65-79.

Krumpal, I., Rauhut, H., Böhr, D. & Naumann, E. (2008) Wie wahrscheinlich ist ‚wahrscheinlich'? Methoden – Daten – Analysen, 2, 1, 3-27.

Leite, W. & Beretvas, N. (2005) Validation of scores on the Marlowe-Crowne Social Desirability Scale and the Balanced Inventory of Desirable Responding. Educational and Psychological Measurement, 65, 1, pp. 140-154.

Lenski, G.E. & Leggett, J.C. (1960) Caste, class and deference in the research interview. American Journal of Sociology, 65, pp. 463-467.

Leverkus-Brüning, R. (1966) Die Meinungslosen. Die Bedeutung der Restkategorie in der empirischen Sozialforschung. Berlin: Duncker & Humbolt.

Liechtenstein, S. & Newman, J.R. (1967) Empirical scaling of common verbal phrases associated with numerical probabilities. Psychonomic Science, 9, 10, pp. 563-564.

Litwin, M.S. (2003) (2nd ed.) How to assess and interpret survey psychometrics. Thousand Oaks, Cal.: Sage publications.

Martin, M.O., Mullis, I.V.S. & Kennedy, A.M. (2007) PIRLS 2006 technical report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Masters, J.R. (1974) The relationship between number of response categories and reliability of Likert-type questionnaires. Journal of Educational Measurement, 11, 1, pp. 49-53.

Matell, M. & Jacoby, J. (1971) Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. Educational and Psychological Measurement, 31, pp. 657-74.

Mayer, H. (2002) Interview und schriftliche Befragung. Entwicklung, Durchführung und Auswertung. München: Oldenbourg Wissenschaftsverlag.

Mittelstaedt, R. (1971) Semantic properties of selected evaluative adjectives: Other evidence. Journal of Marketing Research, 8, 2, pp. 236-237.

Moorman, R.H. & Podsakoff, P.M. (1992) A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organiztaional behavior research. Journal of Occupational and Organizationl Psychology, 65, pp. 131-149.

Mosier, C.I. (1941) A psychometric study of meaning. The Journal of Social Psychology, 13, pp. 123-140.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M. & Foy, P. (2007) PIRLS 2006 international report: IEA's Progress in International Reading Literacy Study in Primary School in 40 Countries. Chestnut Hill, MA: International Association for the Evaluation of Educational Achievement (IEA)

Myers, J. & Warner, G. (1968) Semantic properties of selected evaluation adjectives. Journal of Marketing Research, 5, pp. 409-12.

Nunnally, J.C. & Bernstein, I. (1994) Psychometric theory (3rd ed.). New York: McGraw-Hill.

Oksenberg, L. & Cannell, C. (1977) Some factors underlying the validity of response in self report. Bulletin of the International Statistical Institute, 47, pp. 325-346.

O'Muircheartaigh, C., Krosnick, J. & Helic, A. (2000) Middle alternatives, acquiescence, and the quality of questionnaire data. Unpublished manuscript. Retrieved 19 May 2008 from http://harrisschool.uchicago.edu/about/publications/working-papers/pdf/wp_01_3.pdf.

O'Muircheartaigh, C., Gaskell, G. & Wright, D. (1995) Weighing anchors: Verbal and numeric labels for response scales. Journal of Official Statistics, 11, 3, pp. 295-307.

O'Muircheartaigh, C., Gaskell, G. & Wright, D. (1993) Intensifiers in behavioral frequency questions. Public Opinion Quarterly, 57, 4, pp. 552-565.

Oppenheim, A.N. (1992) Questionnaire design, interviewing and attitude measurement. London: Pinter Publishers.

Oppenheim, A.N. (1966) Questionnaire design and attitude measurement. London: Heinemann.

Paulhus, D.L. (1991) Measurement and control of response bias. In Robinson, J., Shaver, P. & Wrightsman, L. (Eds.), Measures of Personality and Social Psychological Attitudes (pp. 17-59) San Diego, CA: Academic Press.

Paulhus, D.L. (1984) Two-component models of socially desirable responding. Journal of Personality and Social Psychology, 46, pp. 598-609.

Paulhus, D.L. & Van Selst, M. (1990) The spheres of control scale: 10 years of research. Personality and Individual Differences, 11, 10, pp. 1029-1036.

Porst, R. (2000) Praxis der Umfrageforschung. Stuttgart: Teubner.

Presser, S. & Blair, J. (1994) Survey pretesting: Do different methods produce different results? Sociological Methodology, 24, pp. 73-104.

Rockwood, T., Sangster, R. & Dillman, D. (1997) The effects of response categories on questionnaire answers: Context and mode effects. Sociological Methods and Research, 26, 1, pp. 118-40.

Rodgers, W., Andrews, F. & Herzog, R. (1992) Quality of survey measures: A structural modeling approach. Journal of Official Statistics, 8, 3, pp. 251-75.

Schaeffer, N. (1991) Hardly ever or constantly? Group comparisons using vague quantifiers. Public Opinion Quarterly, 55, 3, pp. 395-423.

Scholl, A. (2003) Die Befragung. Socizlwissenschaftliche Methode und kommunikationswissenschaftliche Anwendung. Konstanz: UVK Verlagsgesellschaft.

Schuman, H. & Presser, S. (1996) Questions & answers in attitude surveys. London: Sage Publications.

Schuman, H. & Presser, S. (1978) Attitude measurement and the gun control paradox. Public Opinion Quarterly, 41, pp. 427-39.

Schuman, H. & Presser, S. (1977) Question wording as an independent variable in survey analysis. Sociological Methods and Research, 6, pp. 151-76.

Schwarz, N., Grayson, C.E. & Knäuper, B. (1998) Formal features of rating scales and the interpretation of question meaning. International Journal of Public Opinion Research, 10, 2, pp. 177-183.

Schwarz, N. & Hippler, H. (1991) Response alternatives: The impact of their choice and presentation order. In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, S.Sudman (eds.) Measurement errors in surveys (Chapter 3, pp. 41-56). New York: Wiley-Interscience.

Schwarz, N., Hippler, H., Deutsch, B. & Strack, F. (1985) Response scales: Effects of category range on reported behavior and comparative judgments. Public Opinion Quarterly, 49, 3, pp. 388-95.

Schwarz, N., Knauper, B., Hippler, H., Noelle-Neumann, E. & Clark, L. (1991) Rating scales. Numeric values may change the meaning of scale labels. Public Opinion Quarterly, 55, pp. 570-82.

Simpson, R.H. (1944) The specific meanings of certain terms indicating differing degrees of frequency. Quarterly Journal of Speech, 21, 3, pp. 328-330.

Smith, P.B. (2004) Acquiescent response bias as an aspect of cultural communications style. Journal of Cross-Cultural Psychology, 35, pp. 50-61

Smith, T.W. (2003) Developing comparable questions in cross-national surveys. In J.A. Harkenss, F.J.R. van de Vijver & P.P. Mohler (Eds.), Cross-cultural survey methods (pp. 69-92). Hoboken, N.J.: Wiley Interscience.

Stocké, V. & Hunkler, C. (2007) Measures of desirability beliefs and their validity as indicators for socially desirable responding. Field Methods, 19, 3, pp. 313-336.

Sudman, S. & Bradburn, N.M. (1974) Response effects in surveys. Chicago: Aldine.

Tourangeau, R., Rips, L.J. & Rasinski, K. (2000) The psychology of survey response. New York: Cambridge University Press.

Trometer, R. (1996) Warum sind Befragte "meinungslos"? Kognitive und kommunikative Prozesse im Interview. Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Philosophie der Universität zu Mannheim. Mannheim: Universität Mannheim.

Vidali, J.J. (1975) Context effects on scaled evaluatory adjective meaning. Journal of The Market Research Society, 17, 1, pp. 21-25.

von der Linden, W.J. & Hambleton, R.K. (1997) Handbook of modern item response theory. New York: Springer.

Wedell, D. & Parducci, A. (1988) The category effect in social judgment: Experimental ratings of happiness. Journal of Personality and Social Psychology, 55, pp. 341-56.

Weems, G.H., Onwugbuzie, A.J. & Lustig, D. (2002). Profiles of respondents who respond inconsistently to positively and negatively worded items on rating scales. Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Chattanooga, TN, November 6-8.

Wildt, A.R. & Mazis, M.B. (1978) Determinants of scale response: Label versus position. Journal of Marketing Research, 15, pp. 261-267.

Willis, G.B. (2005) Cognitive interviewing. A tool for improving questionnaire design. Thousand Oaks, Cal.: Sage Publications.

Worcester, R. & Burns, T. (1975) A statistical examination of the relative precision of verbal scales. Journal of the Market Research Society, 17, pp. 181-197.

Wright, D., Gaskell, G. & O'Muircheartaigh, C. (1995) Testing the multiplicative hypothesis of intensifiers. Applied Cognitive Psychology, 9, pp. 167-77.

Wright, B.D. & Masters, G.N. (1982) Rating s cale analysis. Chicago: MESA Press.

Working Paper Series FOR 655

1. Hartmut Kliemt: Priority setting in the age of genomics, December 2007 (1)
2. Marlies Ahlert: If not only numbers count – allocation of equal chances, December 2007 (2)
3. Stefan Felder: The variance of length of stay and the optimal DRG outlier payments, December 2007 (3)
4. Jeannette Winkelhage, Adele Diederich, Simone Heil, Petra Lietz, Felix Schmitz-Justen, Margrit Schreier: Qualitative Stakeholder-Interviews: Entwicklung eines Interviewleitfadens zur Erfassung von Prioritäten in der medizinischen Versorgung, December 2007 (4)
5. Antje Köckeritz: A cooperative bargaining model for two groups of patients, January 2008 (1)
6. Marlies Ahlert and Hartmut Kliemt: Necessary and sufficient conditions to make the numbers count, January 2008 (2)
7. Stefan Felder and Andreas Werblow: Do the age profiles of health care expenditure really steepen over time? New evidence from Swiss Cantons, February 2008 (3)
8. Marlies Ahlert, Wolfgang Granigg, Gertrud Greif-Higer, Hartmut Kliemt, Gerd Otto: Prioritätsänderungen in der Allokation postmortaler Spender-Lebern – Grundsätzliche und aktuelle Fragen, February 2008 (4)
9. Marlies Ahlert, Stefan Felder, Bodo Vogt: How economists and physicians trade off efficiency and equity in medically and neutrally framed allocation problems, February 2008 (5)
10. Adele Diederich, Hartmut Kliemt, Public health care priorities at the polls – a note, March 2008 (6)
11. Stefan Felder: To wait or to pay for medical treatment? Restraining ex-post moral hazard in health insurance, April 2008 (7)
12. Margrit Schreier, Felix Schmitz-Justen, Adele Diederich, Petra Lietz, Jeannette Winkelhage und Simone Heil: Sampling in qualitativen Untersuchungen, Juli 2008 (8)
13. Petra Lietz: Questionnaire design in attitude and opinion research: Current state of an art, September 2008 (9)