

Qualified Answers and Other Doubt Expressions as Indicators of Cognitive Problems in a Health Survey

Stasja Draisma¹, Yfke Ongena², and Wil Dijkstra²
 NESDA, Amsterdam¹
 Free University, Amsterdam²

1. Introduction and Background

If one is interested in response processes in survey interviews, the study of verbal expressions and interactions is indispensable. 'Objective data' obtained in interviews are achieved by the interaction between respondents and interviewers (Maynard and Schaeffer, 2002) and these interactions may reveal problems that respondents might have with questions and difficulties in cognitive processing to arrive at an adequate answer. Consider the following reaction of a respondent to an assertion from a standardized health questionnaire.

1.1 Example 1

I: Would you respond 'yes', 'maybe' or 'no' to the following assertion: 'I am ill'
 R: Well I am chronically ill, but ehm
 R: I can live with that easily
 I: So how would you experience 'I am ill'?
 I: Would you say 'yes' 'maybe' or 'no'?
 R: Well I'd rather answer 'no'.
 R: No, since I do disguise a lot.
 I: o.k.
 I: So 'no' would be the correct alternative?
 R: Yes

This respondents' denial of being ill seems rather odd from the point of view of survey researchers who are interested in sound comparisons of valid answers from large numbers of respondents to questions about such subjective states. The example demonstrates complications in the interpretation process with respect to an evident condition of respondents' life and the way he experiences this condition. Clearly, his interpretation is rather different than those that would be expected from most respondents with such illnesses. Supposedly, the respondent felt the need to justify his answer by the explanation about the disguise of his chronic illness. On the one hand, he seems relatively certain that the label 'ill' does not fit his

condition, on the other hand, he verbalizes some arguments that reveal a difficulty in judging different conflicting types of information (having a chronic disease, against not feeling ill).

A second example illustrates a different kind of cognitive trouble a respondent expresses spontaneously.

1.2 Example 2

I: Would you respond 'yes' or 'no' to the following assertion: 'I am ill'
 [silence]
 R: That's a difficult one, ha ha.
 I: yes
 R: Well as a diabetic... is that considered an illness? That is not really an illness is it?
 I: That's up to your own interpretation.
 R: Let's say 'maybe'.

In this case, the respondent expresses a problem in question interpretation, he is uncertain about the kinds of diseases that are covered by the concept 'illness'. Cognitive problems may occur in different cognitive phases and are sometimes articulated spontaneously.

Such problems may have different reasons: respondents may have difficulties with question interpretation or in mapping a preliminary answer to prescribed answer alternatives, both resulting in *task uncertainty* (Schaeffer and Thomson, 1992). Moreover, they may be unsure about their own true state, *state uncertainty* - either with respect to an opinion asked for or about own behaviors -. State uncertainty may be due to memory problems, or to coming across problems in combining different kinds of information drawn from memory. It can be argued that several verbal expressions can be mapped onto the four steps in cognitive processing of survey questions as were presented by Tourangeau et al. (2000). Example 2 then appears an expression of the first step - question interpretation - after which memory search occurs. The retrieved pieces of relevant

information have to be combined and judged in a subsequent third step, and the problems encountered in example 1 form an expression of complications in processing at this stage. In a last step, the preliminary answer as formulated by the respondent has to be mapped onto the prescribed answer alternatives of the questionnaire. Quite often, respondents come up with a preliminary answer that has to be reformulated (for instance respondents may react with 'true' in the foregoing examples, which has to be changed into 'yes'). Such answers are called 'Mismatch' answers in studies concerning interaction analysis in interviews (Ongena et al., 2004).

An analysis of verbal utterances in a validity study of Draisma & Dijkstra (2004) demonstrated that the occurrence of verbal expressions like doubtwords or hedges (e.g. 'I think', 'I believe', 'possibly', 'maybe') were good indicators of respondents' uncertainty about the correct answer and of measurement error. Schober (2004) used hedges as indicators for misinterpretation of question concepts, but hedges also denote memory and judgment difficulties in health surveys.

For the present paper, verbal interactions in CATI health interviews are analyzed, to identify specific cognitive problems in answering health questions. We will demonstrate that analysis of qualified answers and hedges can be used to detect cognitive problems with health questions.

Moore (2004) calls such qualified answers 'projective reports' and argues they may have two functions: 1. to defer judgment 2. to account for particular answers. In the former case, respondents report a circumstance that appears relevant for formulating a particular answer, and then produce the answer as the consequence of that circumstance.

The latter serves to provide accountability or to reveal the line of reasoning behind the answer. Reporting circumstances in response to a survey question instead of choosing one of the answer options is common practice.

Qualified answers often signify necessary cognitive elaborations (think aloud) to arrive at an answer. A simple qualified answer may be: 'Well I may have six pounds overweight. But yet I would say I am satisfied with my current weight.' More elaborate qualified answers may consist of multiple considerations, and in case of conflicting considerations it takes more time to arrive at an answer, about which the respondent will eventually be less certain. In such cases, it

has been demonstrated that moderate answer alternatives - if available - will be chosen more often (Carabain, 2004).

In the domain of health questionnaires, Mallinson (2002) applied a qualitative analysis of transcripts of verbal utterances to investigate problems in understanding the meaning of question. She identified at least seven different problem categories with well known and validated health measurement instruments. The kinds of problems that hampered the response process were mainly due to the fact that concepts used in the questions were difficult to relate to respondents' own health status or were essentially too general to apply to their situation. For instance, elderly people experience specific physical problems that are difficult to grasp with standard health questions.

In this study, we investigate whether a classification of projective reports can be used to distinguish between verbal reactions on different questiontypes (factual, opinion, frequency estimates of behavior) and what kind of relation exists between such reports - indicating problems - and hedges. It is examined whether different kinds of cognitive problems can be revealed by different expressions. Also, the difference in meaning of expressions before and after the answer will be studied, to examine whether deferral or justification processes occur more often with different question types.

2. Methods

A CATI health questionnaire with 43 questions was used for a sample of 610 respondents. The sample was drawn randomly from a list of Dutch telephone numbers. The Cooperation rate ('COOP 1' or 'Minimum Cooperation Rate 1', AAPOR, 2004) that is, the proportion of interviews of all eligible units ever contacted, amounted to 53.8%. Twelve female interviewers held the interviews.

The questionnaire was based on an evaluation that König-Zahn et al. (1993) performed of health instruments frequently used in the Netherlands, which were often translations of American and English questionnaires. Their evaluation was restricted to instruments assessing general health concepts. Furthermore, they excluded questionnaires that were used in only one study, that were lacking methodological accounts, or that were designed for particular age groups or specific populations. These restrictions essentially fulfilled our requirements to use questions with some external validity to the

health survey domain, and we decided to use some of the questions from the questionnaires they evaluated (for example the General Health Perception Questionnaire (GPHQ) and the Statistics Netherlands Health questionnaire).

In the survey, four different question types occurred:

1. *Own subjective health perceptions*: this type concerns nine assertions about own health and quality of life, such as 'I seem to get ill more easily than other people', 'I accept that I am ill sometimes' derived from the GHPQ scale, as well as two questions concerning the attitude with respect to own weight and one concerning perceived walking speed.

2. *Opinion questions*: Nine questions were posed, of which five concerned the height of health insurance fees for unhealthy behaviors ("People who behave irresponsible in traffic should pay more insurance than people who don't", "Older people should pay more insurance than young people."). Two more questions concerned the role of government in health education and food control.

3. *Questions concerning own behaviors*: Twelve questions concerned background variables like size of the household and car ownership, doctor- and dentist visits, length and weight, but also (un) healthy behaviors like smoking, alcohol consumption and exercising.

4. *Frequency estimates of (un)healthy behavior*: Ten questions were posed about the frequencies of healthy and unhealthy behaviors, such as exercise and walking frequencies, number of days one eats meat and cereals, TV watching, number of alcoholic beverages per week and so on.

The telephone interviews were taped, digitized, transcribed and coded with the use of the computer program SEQUENCE VIEWER (Dijkstra, 2002).

The verbal interactions between interviewer and respondent were divided into separate 'events', that is meaningful separate utterances. Utterances that start with reading a single question until acceptance of an answer – and all utterance in between – make up a single question-answer sequence. A total of 25717 transcribed and coded question answer sequences was usable for analysis.

Hedges in sequences were identified and coded, and projective reports were divided into five different meaningful categories that seemed to cover most different kinds of considerations in the health interview.

The classification scheme contained the following five categories:

1. *Conditionality (accountability / feasibility)*:

This category expresses some kind of constraint with respect to the question topic. Projective reports included in the category can refer to utterances of a general category denoting conditions: 'If I do this then .../ if X happens then.../ if I had done X.... Often, a relation with the attribution of causality exists. 'If a car driver is really careless, he should pay more insurance.' Sometimes, the report is a limitation on a quality, e.g. as a reaction on the assertion 'Getting ill belongs to life', the account 'Well not like this situation I'm in now'.

Also, a comparison with particular situations ('I am a smoker myself, but in a pub, that is something else', 'That depends') and other groups of people or (extreme) behaviors which belong to a category are included. (e.g. 'Well if you get lung cancer and you keep smoking a lot you should pay more'). Feasibility demarcates what is realistic: 'if you have to pay more health insurance when you engage in sports, it is impossible to play soccer anymore.' 'How would the government be able to control that?'

2. *Question interpretation (of concepts and answer alternatives)*:

This category resembles a spontaneous think-aloud concerning interpretation of question elements. Examples are 'Well, then it will refer to people like myself, so I would disagree', "what does 'safe' mean, that depends on how it is interpreted".

3. *Attribute or modification*: indicate verbal reports that refer to own qualities or own behavior. (e.g. 'I Like TV, so I watch more than two hours a day'). Another straightforward example of a reference to respondents' own situation is the following.

I: People have to pay themselves for any extra costs caused by harm due to sport activities

R: Well I had an accident myself, so I would disagree

4. *Uncertainty with respect to own opinion/state*:

Respondents occasionally express uncertainty in a direct verbal manner, as is exemplified with a reaction on a question about food control: "I don't know, I assume the control is done well."

Or: as a reaction to the statement 'Persons who have extra health costs due to smoking should pay themselves': 'Ooh, I really need to think about that. Let's say yes'

5. *Rough estimation / calculation*: This type of qualification refers to numbers, weights and sizes. Sometimes an actual calculation occurs, at

other times, rough verbal estimates like ‘about two or three times a week’ precede or follow an answer. This type of qualified answer only occurred with the frequency estimate questions and the questions concerning own behaviors. The following example illustrates a think aloud calculation:

I: How many hours and minutes do you watch television, on the days that you watch?

R: I watch the news every day for half an hour or so. And then of course the commentaries with regard to the news for half an hour.

R: So, about an hour in total

3. Results

Verbal qualifications either preceded the answers -referring to motives, conditions or present situations - or followed them, serving as justifications. The data used in the analysis consisted of 25717 question-answer sequences of the telephone survey. Sequential analysis is applied to such sequences. The following sequence contain several indicators of uncertainty and deferral.

1. I: Do you agree or disagree with the following statement: *‘People who smoke should pay more health insurance’*

2. R: Ehm
3. R: Well if you get lung cancer and you keep smoking a lot you should pay more
4. R: So I believe, I would agree

Firstly, in the second utterance, a filled pause is given, an indication of uncertainty. The third utterance is an example of a conditional report. Moreover, the respondent uses a hedge (‘I believe’) in expressing the answer.

In seventeen percent of all the sequences, qualifications occur before or after the final answer (of which a little bit more before than after the answer), whereas in sixteen percent of the sequences hedges are uttered. The relationship between hedges and number of reports is significant, both for reports before the answer ($t = 33.4, p < 0.01$) and for reports after the answer ($t = 9.9, p < 0.01$). Since hedges turned out to be such strong indicators for measurement error, one can suspect that sequences with reports signify that problems may also exist with the validity of the responses.

Table 1 shows the number of sequences with qualified answers, and whether these qualifications were given before or after the answer.

Table 1 Percentages of qualifications before and after the answer

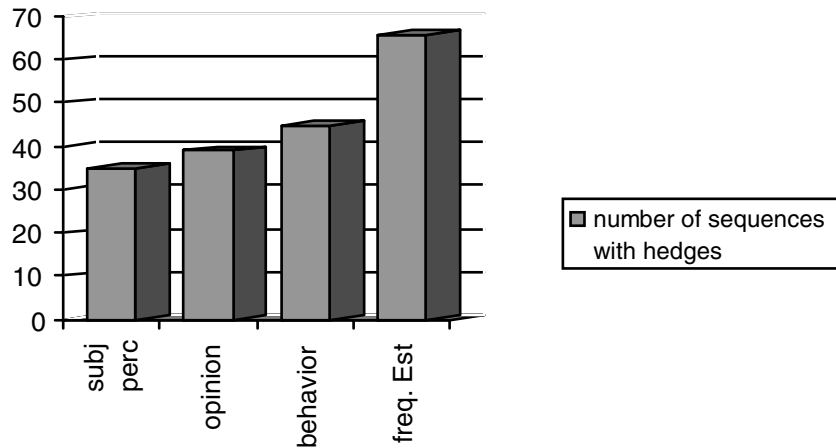
	Subj. perception (N = 7256)	Opinion (N = 5490)	Factual/behavior (N = 7705)	Freq. estimate (N = 5266)	(N = 25717)
Qualification before	5.8 %	9.2 %	6.5%	17.2%	$\chi^2 = 577 p <.001$
Qualification after	6.5%	8.6%	9.5%	9.8%	$\chi^2 = 58 p <.001$

Obviously, questions concerning frequency estimates bring about the largest number of verbal reports. Furthermore, in opinion and frequency estimates, the number of reports before the answer exceed the number afterwards, indicating that for these types more problems occur in the judgment phase, whereas the subjective perception and behavior questions result in more accounts after the answer.

For further analysis (especially the coding of the types of qualifications) 600 sequences were randomly selected, 150 for each of the four question types, and within those 150, 75 with reports before and 75 after the final answer. In these 600 sequences, the percentage of hedges

was 31% and the distribution of hedges over question type is shown in figure 1.

Figure 1 Number of sequences with hedges by each of the four question types (N = 150)



Apparently, expressions of doubt occur more often in factual than in attitudinal questions.

When looking at the kinds of reports given, the most important results are the following. Overall, the following percentages were found for the five types of reports: 1. Conditions: 27%, 2. Interpretation: 8.5 %, 3. Attributes: 45%, 4. Uncertainty: 8.5 %, 5. Calculations: 11 %. Most noteworthy here are the relatively small numbers of reports that signify problems with question interpretation (Phase 1 in Tourangeau's model) and uncertainty (may be corresponding to problems in all four answer phases).

We distinguished reports before and after the final answer, and the most relevant results of this distinction are the following. Firstly, it turns out that (in the small sample) the largest category (44%) of the reports given before the final answer to opinion questions consist of conditionalities (like 'I would like it when people who live healthy pay less insurance, then I would pay less. But no, that's not feasible'). The percentage conditional answers before the behavioral and frequency estimate questions was substantially smaller than for the attitudinal questions (before the answers to subjective belief questions also 40 % of the reports consisted of conditionalities). (see the appendix for all the exact figures).

A substantive category of reports *before* the final answers consist of attributes ('I have false teeth for more than three years, so it must have been longer than three years ago that I visited the

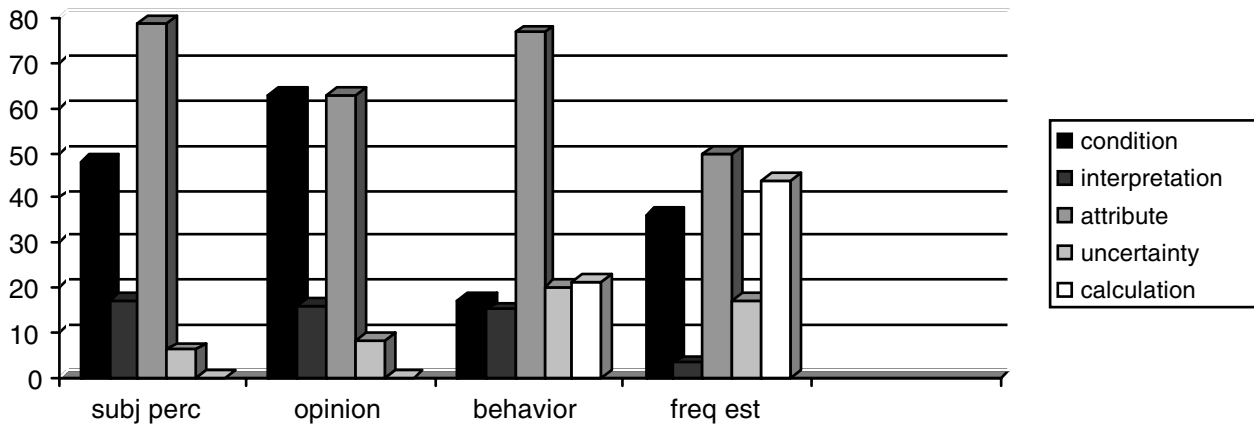
dentist') for three of the four question types, except for the frequency estimate questions. Not surprisingly, the most substantial part of the qualifications for the frequency estimate questions consist of 'calculations and rough estimates' (40 % of the reports), while calculations occur in only 8 % of the factual/behavior questions before the answer and not at all in opinion and subjective estimate questions. Interestingly, more calculation reports are given for behavior questions after the final answer, which may signify that for this question type, such a report may serve as a kind of 'account' (I: do you engage in sportive activities? R: yes. I jog three times a week.)

Also, more expressions of uncertainty (category four) are given before the final answer than after it, consistently for all four types of questions. Yet, substantially more uncertainty reports are given with the behavior and frequency estimate questions than for the attitudinal questions.

Finally, more than half of the reports after the final answers – for all four question types – consist of attributes or modifications: this confirms the expectation that people feel the need to justify their answer with accounts ('I weigh too much. But that is caused by my medicines').

In Figure 2, types of reports by question types are presented for reports given before and after the answers together.

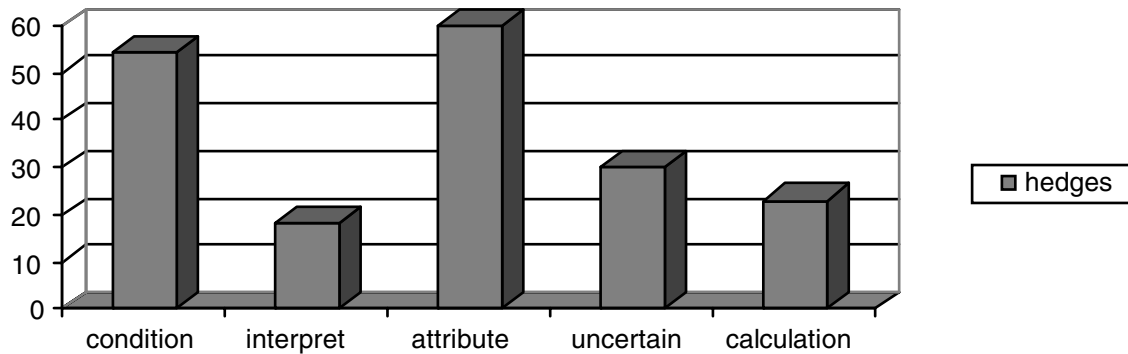
Figure 2 Reports by question type (number per 150 sequences of each question type)



Finally, the distribution of hedges over sequences with the five different types of reports is presented in Figure 3. Most hedges are found

with reports indicating conditionalities and attributes. This is rather straightforward, since we already saw that these two kinds of reports both occur most often with all question types.

Figure 3 Distribution of hedges over types of reports



4. Conclusion

The analysis of verbal expressions in survey interviews does reveal specific cognitive problems with answering questions. Differences in distributions of types of reports over question types signify different problems in cognitive processes: attitudinal questions bring about different kinds and numbers of reports than factual/behavioral questions. Respondents answering opinion questions express more conditionalities than with other types of questions. After a final answer has been provided, most reports – independent of the type of question posed - consist of justifications or

accounts. Before answers are given, other types of reports are present, which are more prominent indicators of deferral of an answer due to for instance difficulties in retrieval and judgment.

The fact that a relation exists between the use of indicators of measurement error (hedges) and the expression of 'projective reporting' produces some doubt with regard to the quality of answers surrounded by reports.

Further analysis of the arguments used in the reports can reveal other kinds of relations: it would be worthwhile to find out whether respondents who express a lot of reports have specific health problems, as expressed in their answers and qualifications. Moreover, an

analysis of numbers and types of reports related to questions with validating information would be more informative with respect to the relation between reports and measurement error. With the present data, it is only possible to investigate whether theoretical relations between health concepts differ for sets of sequences with and without projective reports.

References

Carabain, C. L. (2004). The moderate answer categories: what do respondents actually mean when they choose them? Paper presented at the Annual Conference of the American Association for Public Opinion Research, Phoenix, Arizona, May 13-16.

Draisma, S., & Dijkstra, W. (2004). Response Latency and (Para)Linguistic Expressions as Indicators of Response Error. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 131-149). Hoboken: Wiley.

Dijkstra, W. (2002), *Sequence Viewer 3.0*, Department of Social Research Methods, Vrije Universiteit, Amsterdam.

König-Zahn, C., J.W. Furer, and B. Tax. 1993. *Algemene gezondheid*, vol. 1. Assen: Van Gorcum.

Mallinson, S. (2001). Listening to respondents: a qualitative assessment of the short-form 36 health status questionnaire. *Social Science and Medicine*, 54(11-21).

Maynard, D. W., & Schaeffer, N. C. (2002). Standardization and its discontents. In D. W. Maynard, H. Houtkoop-Steenstra & N. C. Schaeffer (Eds.), *Standardization and Tacit Knowledge. Interaction and Practice in the Survey Interview*. (pp. 3-47). New York: Wiley.

Moore, R. J. (2004). Managing Troubles in Answering Survey Questions: Respondents' Uses of Projective Reporting. *Social Psychology Quarterly*, 67(1), 50-69.

Ongena, Y.P., Dijkstra W. and Draisma, S. (2004) *Conversational and formal questions in Survey Interviews*. Paper presented at Sixth International Conference on Logic and Methodology RC-33, Amsterdam.

Schaeffer, N.C. & Thomson, E. (1992). The Discovery of Grounded Uncertainty; Developing Standardized Questions about Strength of Fertility Motivations. In Marsden, P.V. (Eds), *Sociological Methodology 1992*, 22. pp. 37-82. Oxford: Basil Blackwell.

Schober, M. F., & Bloom, J. E. (2004). Discourse cues that respondents have misunderstood survey questions. *Discourse processes*, 38(3), 287-308.

Tourangeau, R., Rips, L. J. and Rasinsky, K. (2000) *The Psychology of Survey Response*. New York: Cambridge University Press.

Appendix

Type of qualification/report per questiontype

Crosstable of questiontype by type of report

report	questiontype							
	1.bef	1.aft	2.bef	2.aft	3.bef	3.aft	4.bef	4.aft
1	40.00	24.00	44.00	40.00	16.00	6.67	26.67	21.33
2	14.67	8.00	14.67	6.67	16.00	4.00	1.33	2.67
3	37.33	68.00	32.00	52.00	36.00	66.67	14.67	52.00
4	8.00	0.00	9.33	1.33	24.00	2.67	17.33	5.33
5	0.00	0.00	0.00	0.00	8.00	20.00	40.00	18.67
total	75	75	75	75	75	75	75	75

Legend

bef= before aft= after

question type

1 = Subjective perception

2 = Opinion

3 = Behavior

report

1 = Conditionality

2 = Question interpretation

3 = Attribute

4 = frequency estimate

4 = Uncertainty

5 = Calculation