Western Journal of Nursing Research

**Analysis of Quality of Proxy Questions in Health Surveys by
Behavior Coding**

Isabel Benítez, José Luis Padilla and Yfke Ongena

The online version of this article can be found at:

**Additional services and information for *Western Journal of Nursing Research* can be found
at:**

Email Alerts: http://wjn.sagepub.com/cgi/alerts

Subscriptions: http://wjn.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Oct 1, 2012

OnlineFirst Version of Record - Nov 19, 2010

What is This?

Analysis of Quality of Proxy Questions in Health Surveys by Behavior Coding

Isabel Benítez[1], José Luis Padilla[1], and Yfke Ongena[2]

Abstract

The aim of this study is to show how to analyze the quality of questions for proxy informants by means of behavior coding. Proxy questions can undermine survey data quality because of the fact that proxies respond to questions on behalf of other people. Behavior coding can improve questions by analyzing interviewer–respondent interactions. Twenty-nine proxies participated in the pretesting of a disability questionnaire. The questionnaire includes 11 questions related to daily-life limitations as a result of health problems. Interviewer–proxy interactions were coded and analyzed by means of Sequence Viewer program. The percentages, from a methodological perspective, of ideal "question-and-answer" sequences varied from 28% to 76% throughout the 11 questions analyzed. The results obtained pointed out the necessity of reviewing some of the proxy questions analyzed. Behavior coding can improve the quality of proxy questions in health surveys when proxy informants are surveyed.

Keywords

behavior coding, proxy informants, health surveys

[1]University of Granada, Granada, Spain
[2]University of Twente, Enschede, Netherlands

Corresponding Author:
José Luis Padilla, Dpto. Psicología Social y Metodología de las CC. Comportamiento, University of Granada, Campus de Cartuja, 18071 Granada, Spain
Email: jpadilla@ugr.es

The use of indirect informants or "proxies" to obtain information about other household members is common in household health surveys (Duncan et al., 2002; Magaziner, Speaar, Hebel, & Gruber-Baldini, 1996; Pickard et al., 2004; Schwarz & Wellens, 1997). A *proxy* is a person who answers survey questions about the health conditions of other people, whereas *self-reporters* answer about themselves. Proxy reporters are often used to fill in the designated household rosters in household surveys. The answers of the proxy reporter determine the eligibility of other household members to respond to other sections or questionnaires used in the survey. In addition, the use of proxies is frequent in medical processes or disease evaluations. Proxies have been used in an evaluation of the quality of life for reporting on patients with communication difficulties resulting from cerebral injuries (Sneeuw, Aaronson, de Haan, & Limburg, 1997). In turn, proxy and self-reporter answers have been compared to evaluate the validity of a questionnaire for patients who have suffered a stroke (Teixeira-Salmela, Devaraj, & Olney, 2007).

The guidelines and quality criterion for the design of health surveys prepared by the Eurostat Task Force, summing up the present consensus about the use of proxy reporters, indicates that the use of proxy-reporters should be limited only to cases in which (a) people are incapable of responding to questions, due to serious health problems (e.g., dementia, physical or severe mental disability), or (b) to those for whom it is not possible to interview for legal reasons (i.e., minors; Tourangeau, 2003). Nevertheless, using proxies is a common practice in national statistical institutes for the accomplishment of health surveys in numerous countries. The Health Examination Survey database (Koponen & Aromaa, 2001) promoted by the Scientific Institute of Public Health includes information provided by 34 countries from surveys in which proxy informants have been used, including Belgium (Health Interview Survey), Czech Republic (Labour Force Sample Survey), France (Survey on Household Living Conditions), Holland (Continuous Quality of Life Survey), and Spain (Survey of Disability, Personal Autonomy and Dependence Situations). Proxy informants were also used in the National Health Interview Survey on Disability for the National Center for Health Statistics in the United States (Todorov & Kirchner, 2000).

Few studies have looked at how to increase the quality of the answers provided by proxies, in spite of the fact that the use of proxies has been traditionally considered a threat to the quality of survey data (Ávila-Funes, Gray-Donald, & Payette, 2006). One of these studies evaluated the bias in the proxy answers by means of the National Health Interview Survey on Disability carried out in New York. The results showed that proxies used different response strategies than self-reporters (Todorov & Kirchner, 2000). Another study, in which the

proxy answers were evaluated on a scale of cerebral injury impact, showed that proxy and patient evaluations are more consistent when they evaluate observable and specific behaviors, whereas the agreement decreased when the proxy informants made subjective judgments (Duncan et al., 2002). On the other hand, while evaluating the quality of life in patients who have suffered cerebral injuries, it was found that the proxies' evaluations were sensitive to the differences in the patients' functionality (Sneeuw et al., 1997).

Evaluating proxy responses is especially challenging in disability survey contexts because, according to the World Health Organization (WHO, 2009) definition of disability, the classification of a person as having or not having a disability is a subjective judgment as it depends on the interaction between social conventions, individuals, cultural norms, expectations, etc. Therefore, the responses to the questions about whether a person has a disability could vary according to the type of informant (self-reporter vs. proxy), as a result of potential differences between both norms and expectations, but not necessarily as a result of objective information.

Pretest methods can be helpful in improving survey questions. The general objective of pretest methods is the identification of the causes of errors in surveys by means of the analysis of the events occurring during the "question and answer" process (Willis, 2005). Behavior coding is one of the pretest methods used by survey methodologists, either on its own or in combination with other pretest methods such as cognitive interviewing, focus groups, or speech analyses, to optimize the question drafting and the questionnaire design (Presser et al., 2004). In contrast with such pretest methods, behavior coding provides systematic, objective, and replicable results (Groves et al., 2004).

The behavior coding method was developed in the 1960s by Charles Canell to evaluate both the questions and the interviewer behavior (Cannell, Fowler, & Marquis, 1968). Behavior coding is based on the rationale that the interviewer's and respondent's behaviors provide information about potential problems with survey questions related to question phrasing and to questionnaire design by systematically observing the interviewer–respondent interaction (Blair & Srinath, 2008). Moreover, behavior coding allows survey researchers to evaluate the quality of survey questions aimed at specific respondent groups defined by characteristics such as age, educational level, or gender. Nevertheless, little attention has been given to questions designed for respondents with different roles (self-reporter or proxy) in the interview process.

The aim of this study is to show how to analyze the quality of proxy questions by means of behavior coding in a health survey. In this study, the adequacy of the questions to be answered for proxies will be also discussed.

Method

Participants

Twenty-nine proxy informants, 13 men and 16 women, with an age average of 31.06 years took part in the pretest of a disability questionnaire. The educational level of participants was balanced (14 participants with less than 14 years of schooling and 15 participants with more than 14 years of schooling). The sample size of the study is within the interval (15-50) recommended by several authors to maximize the usefulness of results provided by the behavior coding method (Blair & Srinath, 2008).

All participants were Spanish and they provided information only about people with whom they lived and had a direct familiar relationship, for instance, parents, partners, brothers, or sisters. The selection was carried out with regard to various requirements that determine if the participant was eligible, that is to say, they had the same characteristics of the target population of the future health survey in which the tested questions in this study would be administrated.

It was also confirmed that the participants had not previously taken part in a survey pretest. The participants were contacted via associations for disabled person support, and they received 30 euros for taking part in the study.

Materials

The people responsible for carrying out the interviews used interview protocols during the pretest that included demographic questions and 11 "target" questions. The target questions were the ones selected to be analyzed during the pretest by means of behavior coding. These were selected by experts, who evaluated the questionnaire and identified questions that could present difficulties. These experts had a long experience in the field of health surveys and survey methodology. Table 1 shows the 11 questions to be analyzed by means of the behavior coding method.

Procedure

The interviews, in which the questionnaire with the target questions was applied, were conducted by two trained and experienced interviewers (one male and one female). They were specifically instructed to ask target questions as the questions were worded in the questionnaire. The interviews were conducted in a laboratory specially equipped to perform cognitive pretesting. Confidentiality

Table 1. Selected Questions From the Disability Questionnaire

Target Questions

Q.1. Is there any person in your home who has been limited in the performance of habitual activities due to a health problem? The limitation should have lasted or be expected to last more than 1 year.

Q.2. Is there any person in your home who has serious difficulty speaking in an understandable manner and saying meaningful phrases without help?

Q.3. Is there any person in your home who has serious difficulty understanding the meaning of what others say without help?

Q.4. Is there any person in your home who has serious difficulty using the telephone or other devices or means of communication without help and without supervision? Include lip-reading and machines for writing in Braille.

Q.5. As a result of problems of a cognitive or intellectual nature, is there any person in your home who has serious difficulty when intentionally using the senses? For example, paying visual attention, listening attentively, etc.

Q.6. As a result of problems of a cognitive or intellectual nature, is there any person in your home who has serious difficulty learning to read, write, count (or calculate), copy or difficulty learning to use everyday utensils?

Q.7. Is there any person in your home who has serious difficulty moving one's body from one place to another without changing position, without help and without supervision? For example, going from sitting on the bed to sitting on a chair.

Q.8. Is there any person in your home who has serious difficulty changing posture without help and without supervision? For example, getting up from a chair, lying down on the bed, kneeling down, etc. Exclude the action of moving one's body posed in the previous question.

Q.9. Is there any person in your home who has serious difficulty showing other people affection, respect or transmitting feelings, including physical contact such as kisses, caresses, etc.?

Q.10. Is there any person in your home who has serious difficulty forming and maintaining family relationships?

Q.11. Is there any person in your home who has serious difficulty forming and maintaining sentimental or sexual relationships with a partner?

and the exclusive use of the information for research purposes were ensured. Having obtained the respondents' consent, the interviews were audio and video recorded. The interviews were transcribed and two coders used the transcripts and recordings to systematically classify the interviewer and respondent behaviors. The two coders worked independently, and once first classifications were made, they met to analyze discrepancies and reach an agreement.

Table 2. Categories for the Classification of Respondents' Behaviors

Codes	Meaning
During question reading	
Request clarification	Explicit expression for indicating problems in the comprehension of the concepts included in the question or in the task comprehension.
Interruption	The respondent stops the question reading (to request clarification or to answer).
Answer	
Mismatch answer	The response is adequate but is not exactly worded as any of the answer options.
Invalid answer	The response is not related to the question.
Don't know answer	The respondent does not know how to respond.
Qualified answer	The response indicates uncertainty.
Adequate answer	The response fits the objective of the question.

Verbal Behavior Coding

The behavior coding was done by means of the Sequence Viewer program (Dijkstra, 2008). Coders were also trained by experts in Sequence Viewer program. This program provides information about possible problems with the content or the format of the questionnaire, by systematic classification of behaviors occurring during the interview. The analysis begins with the division of the transcripts into sequences. A sequence starts with the reading of a question and ends when the reading of the following question starts (Dijkstra, 1999). The sequences are analyzed by assigning different codes depending on the behaviors occurring during the interviewer–respondent interaction. For example, while interviewers are asking questions, respondents can ask for explanations or extra information (coded as "request for clarification"), and respondents can interrupt the interviewer giving their answers to the question before the interviewer has finished reading or making comments (coded as "interruption"). Answers given by the respondent after interviewers have finished reading the question can be classified in different ways, of which the classification realized by Oksenberg, Cannell, and Kalton (1991) is the most commonly used. This classification has been extended by authors like Van der Zouwen and Smit (2004), Forsyth, Levin, and Fisher (1999), and Ongena (2005). Table 2 shows the coding scheme used in this study, which is primarily based on the classification by Oksenberg et al. (1991).

To evaluate the quality of proxy questions, codes were used in the study as indicators of response accuracy. A scale of accuracy was developed, using extremes represented by the codes "adequate answers" (being the most accurate) and "invalid answers" (being the most inaccurate). The intermediate categories were defined as "mismatch answer," "qualified answer," and "don't know answer."

Depending on the combination of codes assigned, sequences are classified as "paradigmatic sequences," "nonparadigmatic–nonproblematic sequences," and "nonparadigmatic–problematic sequences." A *paradigmatic sequence* is defined as the ideal sequence during the question-and-answer process. An ideal sequence is that in which the delivery of the question is identical to that indicated in the interview protocol, the respondent's answer is adequate, and the interviewer recognizes the answer as being adequate (Ongena & Dijkstra, 2006). A *nonparadigmatic sequence* is problematic or nonproblematic depending on whether the type of behavior occurring is considered to be a problematic influence on the data. In this study, the occurrence of mismatch answers, invalid answers, don't know answers, qualified answers, and requests for clarification all classify the sequence as a problematic sequence. A sequence is classified as nonparadigmatic–nonproblematic when deviations occur that are not problematic (e.g., interruptions).

Sequences are classified considering the codes assigned to each behavior that occurred during the sequence. For example, the occurrence of the behavior "request clarification" causes a sequence to become nonparadigmatic although the respondent's answers were adequate.

Once the sequences were classified, a frequency analysis was performed that consisted first of calculating the frequencies of each type of sequence followed by calculating the rate of the occurrence of problematic answers. When 15% or more of a question's administrations show one or more problematic interactions, it is a widely accepted criterion for determining the question to be flawed (Blair & Srinath, 2008). On the other hand, if the percentage of nonparadigmatic sequences is considered, questions in which the percentage is greater than 60% must be checked (Van der Zouwen & Dijkstra, 2002). Analysis of 319 sequences (i.e., 11 questions × 29 respondents) was conducted using both criteria to illustrate the use of behavior coding in the study.

Results

For the present analysis, 319 sequences (i.e., 11 questions × 29 respondents) were taken into account.

Table 3. Frequencies and Percentages of Each Type of Sequence

	Type of Sequence					
Target Questions	Paradigmatic Sequence		Nonparadigmatic– Nonproblematic Sequence		Nonparadigmatic– Problematic Sequence	
	%	*n*	%	*n*	%	*n*
Q.1. Habitual activities	28	8	21	6	52	15
Q.2. Speak	76	22	3	1	21	6
Q.3. Understand	66	19	24	7	10	3
Q.4. Use the phone	41	12	17	5	41	12
Q.5. Use the senses	66	19	10	3	24	7
Q.6. Learn	69	20	17	5	14	4
Q.7. Move the body	76	22	10	3	14	4
Q.8. Change posture	66	19	14	4	21	6
Q.9. Show affection	76	22	14	4	10	3
Q.10. Family relationships	72	21	10	3	17	5
Q.11. Sentimental relationships	55	16	34	10	10	3

Types of Sequence

First, the behavior coding analyses showed the frequency of occurrence of each type of sequence produced by the proxy informants. Table 3 shows the percentage of occurrence of each type of sequence for each target question.

As Table 3 shows, the percentage of paradigmatic sequences, that is to say, ideal sequences from the methodological point of view, ranges between 28% and 76% for the target questions. The Cramer's V statistic (.2762) indicates a low association between the type of sequence and the target question analyzed. Target Question 1 showed the highest percentage of nonparadigmatic–problematic sequences (52%). Following the usual criteria, Question 1 was recommended for checking, because 72% of the sequences were classified as nonparadigmatic. This high percentage could be due to the content of

Table 4. Frequencies and Percentages of Answer Category Codes

Target Questions	Codes									
	Mismatch Answer		Invalid Answer		Don't Know Answer		Qualified Answer		Adequate Answer	
	%	n	%	n	%	n	%	n	%	n
Q.1. Habitual activities	45	13	0	0	0	0	6	2	90	26
Q.2. Speak	0	0	6	2	0	0	3	1	90	26
Q.3. Understand	0	0	7	2	0	0	0	0	97	28
Q.4. Use phone	14	4	21	6	3	1	7	2	76	22
Q.5. Use senses	7	2	7	2	3	1	7	2	93	27
Q.6. Learn	10	3	3	1	0	0	3	1	100	29
Q.7. Move the body	10	3	3	1	0	0	0	0	100	29
Q.8. Change posture	10	3	10	3	3	1	0	0	90	26
Q.9. Show affection	3	1	7	2	0	0	3	1	93	27
Q.10. Family relationships	3	1	3	1	0	0	0	0	97	28
Q.11. Sentimental relationships	3	1	3	1	3	1	10	3	90	26
Cramer's V	.403		.209		.149		.180		.239	

the question, which is more general and ambiguous than the rest of the target questions.

Codes for Proxy Responses

In this study, we were particularly interested in deviations produced by proxies. Table 4 shows the percentages of adequate and (four types of) inadequate answers for the 11 questions of the disability questionnaire. The percentages per row add up to more than 100% because multiple behaviors can occur in one sequence. For example, the respondent can change the answer once it is coded as "invalid answer."

As Table 4 shows, the Cramer's *V* values reveal a low association between the type of answer produced by the participants and the question analyzed in all the cases except for the code mismatch answer. This code

shows the highest percentage of occurrence for the set of target questions. Question 1 achieved the highest percentage of mismatch answers (45%). The following example represents a situation in which a mismatch answer was produced:

> *Interviewer:* Is there any person in your home who has been limited in the performance of habitual activities due to a health problem? The limitation should have lasted or be expected to last more than 1 year. *Yes, seriously limited; yes, limited but no seriously; not.*
> *Respondent:* Yes

The answer given by the respondent was coded as a "mismatch answer" because it does not fit to any of the response alternatives offered. The high percentage of mismatch answers found in question 1 might be due to respondents' understanding it as a yes/no question without considering the three response alternatives offered.

Question 4 achieved the highest percentage of invalid answers (21%) and the lowest percentage of adequate answers (76%). The following example represents an invalid answer found in Question 4:

> *Interviewer:* Is there any person in your home who has serious difficulty using the telephone or other devices or means of communication without help and without supervision? Include lip-reading and machines for writing in Braille.
> *Respondent:* In my home, nobody knows how to use the machines for writing in Braille.

The answer was coded as an "invalid answer" because its content is not related to the intended objective of the question.

Finally, in Question 11, 10% of answers were registered as qualified. An example from the interviews illustrates the meaning of the code qualified answer.

> *Interviewer:* Is there any person in your home who has serious difficulty forming and maintaining sentimental or sexual relationships with a partner?
> *Respondent:* I don't think so.

In Question 11, the high percentage of qualified answers may indicate that the proxies have doubts when responding to questions on personal topics

Table 5. Frequencies and Percentages of Difficulty Indicator Codes During the Question Reading

	Codes			
	Request Clarification		Interruption	
Target Questions	%	n	%	n
Q.1. Habitual activities	7	2	3	1
Q.2. Speak	3	1	3	1
Q.3. Understand	0	0	3	1
Q.4. Use the phone	14	4	0	0
Q.5. Use the senses	3	1	0	0
Q.6. Learn	3	1	0	0
Q.7. Move the body	0	0	0	0
Q.8. Change posture	7	2	0	0
Q.9. Show affection	3	1	0	0
Q.10. Family relationships	7	2	0	0
Q.11. Sentimental relationships	3	1	0	0
Cramer's *V*	.1740		.1591	

such as sexual or personal relationships. Nevertheless, this leaves the interviewer with a dilemma: Should she further probe for an unqualified answer, or just accept the answer as given? In some cases, this is not necessary, as respondents may spontaneously repair their qualified answer by giving an unqualified adequate answer afterwards.

Difficulty Indicators When Asking Questions

"Request clarification" and "interruption" are codes commonly used in behavior coding as indicators to identify difficulties while interviewers are asking questions. Table 5 shows the frequencies of both codes in the 11 target questions.

As Table 5 shows, the Cramer's *V* values reflect a low association between the behaviors produced by the participants and the target question analyzed. Requests for clarification occurred most frequently with Question 4. Questions 8 and 10 also showed a high percentage in the appearance of this code. The following example demonstrates the occurrence of such a request clarification:

Interviewer: Is there any person in your home who has serious difficulty forming and maintaining family relationships?
Respondent: Family relationships?

Interruptions were coded to some extent in Question 2. The excerpt illustrates an interruption found in Question 2:

Interviewer: Is there any person in your home who has serious difficulty speaking . . .
Respondent: Yes
Interviewer: . . . in an understandable manner and saying meaningful phrases without help?

In this specific case, difficulties could arise from an interruption, since the respondent is answering the question before hearing all the elements that have to be considered (Van der Zouwen & Dijkstra, 2002).

Discussion

The aim of the study was to illustrate how to analyze the quality of the questions intended for proxy respondents in a health survey by means of behavior coding. The results from the behavior coding application to the disability questionnaire pretested in the study allowed the quality of the proxy questions to be analyzed.

The general results showed percentages of paradigmatic sequences between 28% and 76% for the set of 11 target questions. Only Question 1, habitual activities, achieved more than 60% of nonparadigmatic sequences.

The results highlighted some questions to be checked or in which it was necessary to examine the proxies' behavior in detail. These problems might be due to the characteristics of the questions, or to the role represented by the informants. For example, Question 1 (habitual activities) is worded as a yes/no question while three alternatives are offered to the respondent. This is a problem that is common in survey questionnaire design (Ongena, 2003). In addition, two of the three options are positive ("Yes, seriously limited" and "Yes, limited but no seriously"), and one is negative ("Not"). Thus, researchers find an adequate answer in cases in which the respondents' answer is negative, but a large percentage of mismatch answers when the respondents' answers are positive but they replied with a simple yes. Assessing how serious the limitation was and distinguishing between the affirmative alternatives can be a difficult task for proxies. On the other hand, proxy behavior

could cause measurement error because either proxies focus on aspects that are not the aim of the question (Question 4 on the use of the telephone) or they face nonobservable or sensitive topics (Question 10 on family relationships). Possible impact of demographics, such as educational level, degree of family relationship, and so on, on proxy questions was not specifically addressed in our study because of its particular design, which can be considered a limitation.

When using proxies, survey researchers consider several factors. The difficulty of the task and the motivation for responding to questions could be different for self-reporters than for proxies. In addition, proxy respondents may have less information available in their episodic memory (Schwarz & Wellens, 1997). More studies focused on comparing the proxies and self-reporter behavior are necessary, as well as evaluating the convergence between the answers provided by both types of informants. Future research may address these topics.

Respondent behavior can be studied from multiple perspectives, including more qualitatively oriented studies. For example, Collins, Shattell, and Thomas (2005) address how to deal with potentially problematic interviewee behaviors, such as flattery, filtration, or statements indicative of social desirability response bias for qualitative research. Behavior coding as a method provides a systematic approach to analyze interviewer and respondent behavior, is flexible, and offers the possibility of obtaining qualitative and quantitative information that help survey methodologists improve survey data quality. In comparison with other pretest methods, behavior coding is focused on the participant's behavior. The assumption behind behavior coding is that the interviewer–respondent interaction can provide very useful information about potential problems with question phrasing and questionnaire design. This information allows survey researchers to identify questions with a high percentage of "problematic behaviors" as questions that should be revised.

Behavior coding also presents some limitations. For example, it is possible that a respondent gives an adequate answer although he has not understood the real sense of the question. In fact, there may be a gap between respondents' observed behaviors and their understanding of the key concepts in the questions. Combining behavior coding and cognitive interviewing can resolve that gap. Future research in the pretest methods field should address how to combine evidence provided by different pretest methods.

On the other hand, it is necessary to reach a greater consensus about the criteria used to check the questions on the basis of the results obtained by means of the behavior coding. A review of studies in which behavior coding was used found that some authors consider those questions problematic in which the percentage of adequate answers was lower than 85% and some

other authors when that was lower than 90%, whereas others focused on the percentage of inadequate answers, recommending to review the questions in which the percentage is greater than 15% (Van der Zouwen & Smit, 2004). The criteria used can cause changes in the conclusions drawn because, for example, an adequate answer can occur after an inadequate answer. If an inadequate answer criterion is used, a question can be eliminated although a high percentage of final adequate answers has been reached.

Behavior coding has shown its usefulness in evaluating the quality of questions designed for proxy informants by providing detailed information about the participants' behavior and facilitating the detection of possible sources of measurement error. However, more research is needed to find out the causes of question problems identified by coding behavior and their consequences when results of behavior coding studies are applied in survey questionnaire design, especially when proxy questions are included in the survey questionnaire. Nevertheless, as Oksenberg et al. (1991) highlight, there is convincing evidence of the usefulness of behavior coding in improving the quality of survey questions providing quantitative, systematic, and replicable results.

Declaration of Conflicting Interests

Funding

References

Ávila-Funes, J. A., Gray-Donald, K., & Payette, H. (2006). Medición de las capacidades físicas de adultos mayores de Quebec: un análisis secundario del estudio NuAge [Measurement of physical capacities in the elderly: A secondary analysis of the Quebec longitudinal study NuAge]. *Salud Pública en México, 48*, 446-454.

Blair, J., & Srinath, K. P. (2008). A note on sample size for behavior coding pretests. *Field Methods, 20*, 85-95.

Cannell, C. F., Fowler, F. J., & Marquis, K. H. (1968). The influence of interviewer and respondent psychological and behavioral variables on the reporting of household interviews. *Vital and Health Statistics, 2*(26), 1-65.

Collins, M., Shattell, M., & Thomas, S. P. (2005). Problematic interviewee behaviors in qualitative research. *Western Journal of Nursing Research, 27*, 188-199.

Dijkstra, W. (1999). A new method for studying verbal interactions in survey interviews. *Journal of Official Statistics, 15*, 67-85.

Dijkstra, W. (2008). Sequence viewer (version 4.4a) [Computer software]. Free University of Amsterdam, Netherlands.

Duncan, P., Min Lai, S., Tyler, D., Perera, S., Reker, D. M., & Studenski, S. (2002). Evaluation of proxy responses to the stroke impact scale. *Stroke, 33*, 2593-2599.

Forsyth, B., Levin, K., & Fisher, S. (1999). Test of an appraisal method for establishment survey questionnaires. In *Proceedings of the ASA Section on Survey Research Methods* (pp. 145-149). Alexandria, VA: American Statistical Association.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Hoboken, NJ: Wiley.

Koponen, P., & Aromaa, A. (2001). Health examination surveys (HES) in the European Union: Review of literature and inventory of surveys in the EU/EFTA Member States. Retrieved from the Scientific Institute of Public Health website http://www.iph.fgov.be/keywords.asp?Lang=EN&ReportID=2107

Magaziner, J., Speaar, S., Hebel, J. R., & Gruber-Baldini, A. (1996). Use of "proxies" to measure health and functional status in epidemiologic studies of community-dwelling women. *American Journal of Epidemiology, 143*, 283-292.

Oksenberg, L., Cannell, C., & Kalton, G. (1991). New strategies for pretesting survey questions. *Journal of Official Statistics, 7*, 349-365.

Ongena, Y. P. (2003, June). *Pre-testing the ESS-questionnaire using interaction analysis*. Paper presented at the European Social Survey CCT meeting, Sociaal en Cultureel Planbureau, The Hague, Netherlands.

Ongena, Y. P. (2005). *Interviewer and respondent interaction in survey interviews* (Unpublished doctoral dissertation). Vrije Universiteit, Amsterdam, Netherlands.

Ongena, Y. P., & Dijkstra, W. (2006). Methods of behavior coding of survey interviews. *Journal of Official Statistics, 22*, 419-451.

Pickard, A. S., Johnson, J. A., Feeny, D. H., Ashfaq, M. D., Carriere, K. C., & Abdul, M. N. (2004). Agreement between patient and proxy assessments of health-related quality of life after stroke using the EQ-5D and health utilities index. *Stroke, 35*, 607-612.

Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., & Singer, E. (2004). *Methods for testing and evaluating survey questionnaires*. New York, NY: Wiley-Interscience.

Schwarz, N., & Wellens, T. (1997). Cognitive dynamics of proxy responding: The diverging perspectives of actors and observers. *Journal of Official Statistics, 13*, 159-174.

Sneeuw, K. C., Aaronson, N. K., de Haan, R. J., & Limburg, M. (1997). Assessing quality of life after stroke: The value and limitations of proxy ratings. *Stroke, 28*, 1541-1549.

Teixeira-Salmela, L. F., Devaraj, R., & Olney, S. J. (2007). Validation of the human activity profile in stroke: A comparison of observed, proxy and self-reported scores. *Disability Rehabilitation, 29*, 1518-1524.

Todorov, A., & Kirchner, C. (2000). Bias in "proxies." Reports of disability: Data from the National Health Interview Survey on Disability. *American Journal of Public Health, 90*, 1248-1253.

Tourangeau, R. (2003). Cognitive aspects of survey measurement and mismeasurement. *International Journal of Public Opinion Research, 15*, 3-7.

Van der Zouwen, J., & Dijkstra, W. (2002). Testing questionnaires using interaction coding. In D. Maynard, H. Houtkoop-Steenstra, N. Schaeffer, & J. Van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the Survey interview* (pp. 427-447). New York, NY: John Wiley.

Van der Zouwen, J., & Smit, J. H. (2004). Evaluating survey questions by analyzing patterns of behaviour codes and transcripts of question-answer sequences: A diagnostic approach. In S. Presser, J. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 109-130). New York, NY: John Wiley.

Willis, G. B. (2005). *Cognitive interviewing*. Thousand Oaks, CA: Sage.

World Health Organization. (2009). *ICIDH-2: International Classification of Functioning, Disability and Health (ICF)*. Geneva, Switzerland: Author.