



Are interactional behaviors exhibited when the self-reported health question is asked associated with health status?

Dana Garbarski ^{a,b,*}, Nora Cate Schaeffer ^{a,c}, Jennifer Dykema ^c

^a Department of Sociology, University of Wisconsin-Madison, 8128 William H. Sewell Social Science Building, 1180 Observatory Drive, Madison, WI 53706, United States

^b Department of Population Health Sciences, University of Wisconsin-Madison, 707 WARF Building, 610 N. Walnut St., Madison, WI 53726, United States

^c University of Wisconsin Survey Center, 475 N. Charter St., Madison, WI 53706, United States

ARTICLE INFO

Article history:

Received 8 December 2010

Available online 28 April 2011

Keywords:

Self-reported health

Interactional behaviors

Interaction coding

Interviewer–respondent interaction

Cognitive ability

Health

Question–answer sequences

ABSTRACT

The self-reported health question summarizes information about health status across several domains of health and is widely used to measure health because it predicts mortality well. We examine whether interactional behaviors produced by respondents and interviewers during the self-reported health question–answer sequence reflect complexities in the respondent's health history. We observed more problematic interactional behaviors during question–answer sequences in which respondents reported worse health. Furthermore, these behaviors were more likely to occur when there were inconsistencies in the respondent's health history, even after controlling for the respondent's answer to the self-reported health question, cognitive ability, and sociodemographic characteristics. We also found that among respondents who reported “excellent” health, and to a lesser extent among those who reported their health was “very good,” problematic interactional behaviors were associated with health inconsistencies. Overall, we find evidence that the interactional behaviors exhibited during the question–answer sequence are associated with the respondent's health status.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

The self-reported health question – e.g., “Would you say your health in general is excellent, very good, good, fair, or poor?” – summarizes information about health across several domains and is widely used to measure health status because of its ability to predict morbidity and mortality (Idler and Benyamini, 1997). Researchers have demonstrated that self-reported health is related to multiple domains of health including illnesses, symptoms of undiagnosed diseases, judgments about the severity of illness, family history, dynamic health trajectory, health behaviors, and the presence or absence of resources for good health (Idler and Benyamini, 1997). In sum, “a very long list of variables is required to explain the effect of one brief 4- or 5-point scale item...” (Idler and Benyamini, 1997, p. 31). We seek to demonstrate that there is additional health information to be gleaned from the self-reported health question; in particular, that information from the interviewer–respondent interaction during the self-reported health question–answer sequence may capture information about respondents' health status beyond that provided solely by their answer to the self-reported health question.

* Corresponding author at: Department of Sociology, University of Wisconsin-Madison, 8128 Sewell Social Science Building, 1180 Observatory Drive, Madison, WI 53706, United States. Fax: +1 608 262-8400.

E-mail address: dgarbars@ssc.wisc.edu (D. Garbarski).

2. Theory and evidence

2.1. Dimensions of health associated with the self-reported health question

Two broad sets of studies have investigated the dimensions of health respondents consider when they answer the self-reported health question. First are studies that investigate the associations between self-reported health and other measures of health to determine which of the measures are more strongly associated with self-reported health. An inference is then made that the measures that are more strongly associated with self-reported health were weighed more heavily by respondents when constructing their answers. These studies have found the following: current health experience is more strongly associated with self-reported health than is prior health experience (Benyamini et al., 1999); indicators of positive health are as important in determining future and current self-reported health as negative indicators (Benyamini et al., 2000); and men's self-reported health is associated with serious, life-threatening diseases while women's self-reported health is associated with both life-threatening and non-life-threatening diseases (Benyamini et al., 2000). In addition, the dimensions of health that respondents rate as important for their self-reported health vary by the response option (e.g., "excellent," "very good," "good," "fair," or "poor") they select. For example, respondents who selected "poor" or "fair" rated current disease status as important, while those choosing "excellent," "very good," and "good" rated risk factors and positive indicators as important; respondents in all categories rated overall functioning and vitality factors as important (Benyamini et al., 2003).

Other studies have used follow-up probes to ask respondents what they were thinking about when they answered the self-reported health question in order to ascertain how respondents constructed their answers. For example, Groves et al. (1992) found that respondents who used external cues (absence or presence of illness, health service use, and outcome of physical exam) were slightly less likely than those who used internal cues (feelings, physical performance or ability, affect) to report "excellent," "fair," or "poor" health. Further, those who reported that they considered their health in more recent time were less likely to report "excellent" and more likely to report "good" or "fair" health. Krause and Jay (1994) reported that overall there was not a significant relationship between self-reported health and the content of respondents' reports to follow-up probes. However, interesting patterns emerged in the data: respondents who reported that they compared their health to that of others were more likely to select "excellent," while respondents whose health referents were physical functioning, health problems, or health behaviors were more likely to select "good." In cognitive testing of the self-reported health question, Canfield and colleagues (2003) reported that respondents considered a variety of situational factors in constructing an answer; these included reporting "good" health despite a long list of serious health conditions, weighing how well a condition was controlled by medication, comparing themselves with others their own age or with similar medical conditions, or considering a prior question about physical limitations.

We expect that interactional behaviors exhibited during the self-reported health question–answer sequence may provide an indirect and unobtrusive way to capture information about the respondent's health status beyond that provided by the answer to the self-reported health question. In contrast to the approaches outlined above, coding features of the interviewer–respondent interaction does not require additional health-related questions or follow-up probes. We seek to show that behaviors exhibited during the self-reported health question–answer sequence may provide more information about the respondent's health than the answer to the question alone. This information may be useful when additional questions or follow-up probes are not available.

2.2. Interactional behaviors and model of the response process

Interaction behavior coding has been used as technique to pre-test survey instruments. Behavior coding identifies questions as problematic when the interaction between the respondent and interviewer deviates from the paradigmatic question–answer sequence of standardized interviewing (Schaeffer and Maynard, 1996; see van der Zouwen and Smit, 2004, for a review; see discussion below). Some interactional behaviors – such as response latency and expressions of uncertainty by respondents, and probing of respondents by interviewers – have been shown to be associated with inaccurate or unreliable answers, the difficulty of the task, and cognitive ability (Draisma and Dijkstra, 2004; Dykema et al., 1997; Fowler and Cannell, 1996; Hess et al., 1999; Knäuper et al., 1997; Mathiowetz, 1999; Schaeffer and Dykema, 2004; Schaeffer et al., 2008; van der Zouwen and Smit, 2004).

Researchers have begun to link behavior coding to the cognitive processes respondents follow to answer survey questions, because some behaviors of respondents can be viewed as by-products of the information processing that occurs when a respondent answers a survey question (Fowler and Cannell, 1996; Groves, 1996; Holbrook et al., 2006). In constructing an answer to a survey question, a respondent may progress through four stages: comprehension of the question, retrieval of relevant information from memory to answer the question, use of retrieved information to make judgments, and selection and reporting of an answer (Tourangeau et al., 2000). The actual stages of cognitive processing of the response can vary depending on the wording of the question, how motivated the respondent is to be accurate, and the respondent's cognitive ability.

We propose that the interactional behaviors exhibited by respondents and interviewers during the self-reported health question–answer sequence may also be related to the content of what the question is asking, that is, with various dimensions

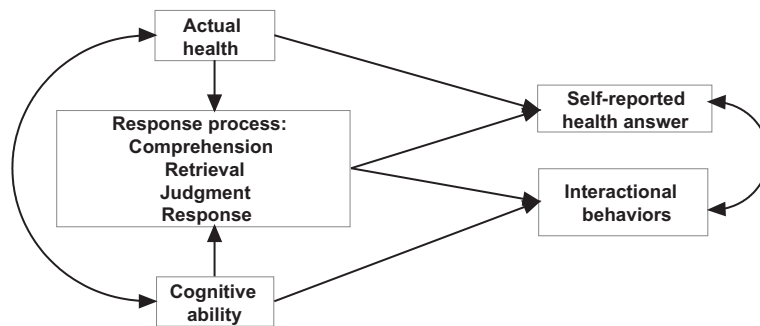


Fig. 1. Conceptual model of the response process linking respondent's health status to interactional behaviors during the self-reported health question-answer sequence.

of health. More specifically, we expect that when the respondent's health history is complex, inconsistencies between health conditions and functioning may lead to challenges in cognitive processing that in turn may be expressed in behaviors the respondent and interviewer exhibit during interaction. As a result, these behaviors may provide information about health status that may be particularly useful when limited data on health are collected.

2.3. Conceptual model of self-reported health question-answer sequence

Our conceptual model is depicted in Fig. 1. The diagram indicates that the response process is affected by the respondent's actual health and her cognitive ability; actual health and cognitive ability are also correlated. The respondent's actual health and her response process affect her answer to the self-reported health question, and her response process and cognitive ability determine which behaviors are exhibited during the interaction. Finally, the respondent's self-reported health answer and interactional behaviors are associated because they are co-produced by the respondent after the question is posed by the interviewer (see also Schaeffer and Dykema, 2011).

Prior research has shown that some interactional behaviors, such as response latency, are associated with cognitive ability as measured by performance on cognitive assessments (e.g., Schaeffer et al., 2008). For self-reported health, our model suggests that this association could include both the general effect of cognitive ability on behaviors, as well as the effect that cognitive ability has on the cognitive processing of this particular question. We also expect that the respondent's actual health will influence behaviors during the interaction in three ways: by implicating a specific answer to the self-reported health question, through the effects of actual health on cognitive ability, and through the effect of the respondent's actual health on the processing of the question.

More specifically, in this paper we focus on how inconsistencies in respondents' reports of disease and functioning might be associated with behaviors during the self-reported health question-answer sequence. The complexity of the information respondents have about their actual health may require them to combine disparate types of information, such as disease and functioning, which may make it more difficult to integrate relevant information, formulate judgments about health, and map an answer onto the response options. We seek to show that respondents with more complex health histories are more likely to exhibit certain interactional behaviors thereby providing evidence that these behaviors may indicate difficulties in the response process (e.g., difficulties in combining information about disease and functioning or difficulties in mapping a resulting judgment onto one of the offered options). As a result, behaviors exhibited during the self-reported health question-answer sequence may provide additional information about health status.

3. Research hypotheses

We examine a subset of interactional behaviors produced when interviewers administer and respondents answer the self-reported health question. We select behaviors previously identified as indicating potential problems in the response process, which we refer to as problematic interactional behaviors. These behaviors include: tokens (such as "uh"), expressions of uncertainty, and long response latencies produced by respondents; pre-emptive and follow-up behaviors by interviewers; and nonparadigmatic question-answer sequences and multiple exchange levels resulting from the interviewer-respondent interaction (definitions provided in discussion below). Because these behaviors are co-produced when respondents answer the self-report health question, they vary along with responses to the self-reported health question. We predict that:

Hypothesis 1. Problematic interactional behaviors will be more likely to occur and response latencies will be longer when respondents report worse self-reported health, because less healthy respondents have a more complex response task and may be more likely to have cognitive abilities impaired by poor health.

Hypothesis 2. Problematic interactional behaviors will be more likely to occur and response latencies will be longer when respondents have inconsistent health statuses (i.e., inconsistencies in disease and functioning), even after controlling for the respondent's answer to the self-reported health question, a measure of cognitive ability, and sociodemographic control variables. The presence of the predicted relationship indicates that the behaviors may provide information about the respondent's health status beyond that provided by the respondent's answer to the self-reported health question and other potential confounders.

Hypothesis 3. There will be a positive correlation between the occurrence of problematic interactional behaviors and a measure of inconsistent health status among respondents who answer the self-reported health question using the same response option (e.g., among respondents who select “excellent”). The presence of the predicted relationship suggests that the behaviors carry information about differences in health status among respondents with the same self-reported health answer.

4. Methods

4.1. Data

4.1.1. Sample selection

Data for this study are provided by the 2004 telephone administration of the [Wisconsin Longitudinal Study \(WLS\)](#), a longitudinal study of 10,317 randomly selected respondents who graduated from Wisconsin high schools in 1957. While the WLS sample is homogenous with respect to race (white), region (grew up in Wisconsin), and education (high school graduate and above), the strength of the study lies in the variety and depth of topics measured at several points across the life course, such as educational plans, occupational aspirations, social influences, and, more recently, physical and mental health status ([Hauser, 2009; Sewell et al., 2004](#)).

We analyze interviewer–respondent interaction produced during the administration of the self-reported health question and several other questions about physical and mental health. Our analytic sample of 355 cases was randomly selected as follows. A total of 230 interviewers participated in the 2004 wave of the WLS. We randomly sampled 100 interviewers from the 137 interviewers who completed 4 or more interviews during even replicates of the survey.¹ The total number of interviews completed by our random sample of 100 interviewers ranged from 4 to 147 (mean 49.95, standard deviation 35.39). Due to budget constraints, we then sampled between 3 and 5 respondents from each of the sampled interviewers (mean 4.49, standard deviation 0.78).

To sample respondents within interviewers, we first stratified by the respondent's cognitive ability, measured by the respondent's high school IQ score normalized for the entire WLS sample. Stratification by IQ was performed by splitting the IQ scores of WLS respondents into low, medium, and high tertiles. Next, within interviewer, we randomly selected up to 2 respondents with low cognitive ability, 2 respondents with high cognitive ability, and 1 respondent with medium cognitive ability. Cognitive ability was associated with overall participation in the WLS ([Hauser, 2005](#)), with the consequence that the WLS sample became more homogenous over time with respect to the IQ scores obtained in 1957. Stratifying by cognitive ability ensures that we have an adequate number of respondents with lower measures of cognitive ability in the sample in order to determine whether cognitive ability is associated with the respondent's interactional behaviors.

[Table 1](#) shows that the mean and standard deviation of IQ for our analytic subsample are similar to those measured for the entire sample in 1957. Thus, stratifying by IQ led to an adequate representation of respondents with lower cognitive ability in our analytic subsample. In general, our analytic subsample is overwhelmingly similar to the entire WLS sample across a range of sociodemographic characteristics ([Table 1](#)), and these similarities remain when both samples are stratified by IQ score (not shown).

4.1.2. Coding process

We examine the interviewer–respondent interaction during the self-reported health question–answer sequence, beginning with the interviewer's administration of the question and ending with the respondent's final answer. The interaction was coded using an elaborate scheme that segmented the interaction into a series of events and assigned codes to specific behaviors, a subset of which is described below.

Five former WLS interviewers were hired to code the interviewer–respondent interactions. Coders went through a two-day training program developed by the authors. Training involved a detailed review of transcription and coding manuals which were populated with definitions, examples, and practice cases. Coders were then assigned to transcribe and code 7 practice cases – ranging from easy to difficult – in their entirety and on their own. Coders met with the authors to review

¹ We required that interviewers had to have completed a minimum of four interviews in order to ensure there were a sufficient number of respondents per interviewer for analysis. The WLS sample is divided into random replicates based on gender, socioeconomic status, and mental ability ([Hauser, 2005](#)). Cases for our analytic sample were drawn from the even-numbered replicates in order to spread our analytic sample across the life of the project. We set aside cases from the odd-numbered replicates for future analyses.

Table 1

Characteristics of Wisconsin Longitudinal Study and analytic samples in 2004.

Sample characteristics in 2004	Wisconsin Longitudinal Study sample in 2004			Analytic subsample		
	N	Mean or %	SD	N	Mean or %	SD
Sex (female = 1) ^a	10317	51.6%		355	52.1%	
Residence (WI = 1)	7265	67.5%		355	63.9%	
Marital status (married = 1)	7692	78.1%		355	78.3%	
Years of education	7264			355		
High school		56.7%			53.0%	
Some college		16.0%			14.9%	
Bachelor's		13.5%			15.2%	
Post-bachelor's		13.8%			16.9%	
Current employment	6613			321		
Employed, never retired		34.2%			34.0%	
Employed, retired before		14.8%			14.6%	
Not employed, never retired		4.1%			4.1%	
Not employed, retired before		47.0%			47.4%	
Total household income	7186	68686.4	154590.3	355	68288.5	100480.2
Self-reported health	7719			355		
Excellent		23.9%			28.7% ^b	
Very good		38.0%			35.2%	
Good		28.0%			26.8%	
Fair		7.9%			6.8%	
Poor		2.3%			2.5%	
Health utilities index (Mark 3)	7262	0.8	0.2	355	0.8	0.2
Reported health conditions (0-5)	6323	1.2	1.0	355	1.4 ^b	1.1
IQ ^a	10317	100.5	14.9	355	102.0	18.0

^a Measured in 1957 wave of the Wisconsin Longitudinal Study.^b Significantly different from entire Wisconsin Longitudinal Study sample, $p < .05$.

the practice cases and correct mistakes. During production coding, coders were encouraged to submit their questions about the coding of the cases, and the authors disseminated answers to all coders so that everyone was privy to the same information regarding transcription and coding conventions. Because this analysis is part of a larger project that coded the entire health section (not just the self-reported health item) as well as two cognitive tasks, the coding phase of the project lasted 16 months. We started the project with 5 coders and ended with 3 coders.

Sequence Viewer (Wil Dijkstra, <http://www.sequenceviewer.nl/>) was used to code events in the interviewer–respondent interactions. Coders used a sound editing program called Audacity to listen to the sound files and to time the response latencies. Audacity has a visual representation of the sound waves called a spectrogram, which displays the sound frequencies of the interviewer–respondent interactions. In order to time the response latencies in tenths of seconds, coders were able to highlight on the spectrogram the length of time in tenths of seconds from the end of the interviewer's administration of the item to the beginning of the first codable answer from the respondent. Our coders were able to be more precise using this method of timing response latency than with a method relying on reaction time, such as manual manipulations of onset and offset timings.

4.1.3. Reliability

In order to assess inter-coder reliability of our coded interactions, a sample of 30 cases was independently double-coded by 5 coders and a measure of inter-rater agreement, Cohen's Kappa, was produced. Cohen's Kappa is a statistical measure of inter-rater agreement that is generally thought to be a more robust measure than percent agreement since it takes into account the agreement between coders that may occur by chance (Cohen, 1960).

The formula for Cohen's Kappa is $\kappa = (\text{Pr}(a) - \text{Pr}(e)) / (1 - \text{Pr}(e))$, where $\text{Pr}(a)$ is the relative observed agreement among coders, and $\text{Pr}(e)$ is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each coder randomly coding each category (Cohen, 1960). To calculate a Kappa per code, the observed agreement for a particular code is the number of times both coders assign a particular code to the same event plus the number of times both coders do not assign a particular code to the same event, divided by the total number of events, or $\text{Pr}(a) = (A + D) / I$ in Table 2. The hypothetical probability of chance agreement for a particular code is calculated by summing the probability that both coders use code X with the probability that both coders do not use code X, or $\text{Pr}(e) = (E/I) * (G/I) + (F/I) * (H/I)$ in Table 2. In order to calculate an overall Kappa, the observed agreement is calculated for each code and then summed across all the codes. The same is done for the probability of chance agreement, and then the summed values of $\text{Pr}(a)$ and $\text{Pr}(e)$ are used in the formula to calculate an overall Kappa for all codes used in the analysis.

According to Fleiss (1981), values of Kappa greater than .75 indicate excellent agreement beyond chance, values between .40 and .75 indicate fair to good agreement, and values below .40 indicate poor agreement. Using Sequence Viewer, we calculated Cohen's Kappa for all of the coded events in the self-reported health question–answer sequence at .924. Because we coded for several interactional behaviors as part of our larger project, our overall Kappa for the self-reported health item of

Table 2

Frequency of agreement between coder 1 and coder 2 about the assignment of code X, used for computing Kappa.

		Coder 1		Total
		Code X assigned	Code X not assigned	
Coder 2	Code X assigned	Cell A: Number of events coder 1 and coder 2 assigned code X	Cell B: Number of events coder 1 did not assign code X but coder 2 did	Cell G: Total number of events coder 2 assigned code X
	Code X not assigned	Cell C: Number of events coder 1 assigned code X but coder 2 did not	Cell D: Number of events coder 1 and coder 2 did not assign code X	Cell H: Total number of events coder 2 did not assign code X
	Total	Cell E: Total number of events coder 1 assigned code X	Cell F: Total number of events coder 1 did not assign code X	Cell I: Total number of events coded for analysis

.924 includes behaviors that were not used in the current analysis. Furthermore, Kappas for individual codes vary around the overall Kappa of .924, and Kappas for rarer codes may be unstable because only 30 cases were double-coded for reliability analysis.²

4.2. Measures

4.2.1. Dependent variables

We examine six problematic interactional behaviors and patterns that may indicate difficulties in the response process, due to either the complexity of the respondent's actual health or to her cognitive ability.³ It should be noted that while each measure in the analysis represents a distinct interactional construct of interest, some specific behaviors are used in the operationalization of more than one construct.

Tokens such as “um,” “uh,” or “well” are sometimes labeled “disfluencies” and have been interpreted as indicating disruption in the speaker's cognitive processing (e.g., Bortfeld et al., 2001). Some tokens have also been described as “continuers,” because they appear to respond to a prior utterance in a way that signals to the speaker to continue (Schegloff, 1981). We include an indicator for whether the respondent uttered any tokens (0 = none, 1 = 1 or more).

Respondents express uncertainty in many ways. We include an indicator that is coded “1” (versus “0” otherwise) if the respondent exhibited at least one of the following behaviors: a report or consideration, in which the respondent provides information that is either stated as an answer or offered as an explanation for an answer (e.g., “my mental health is ok, my physical health is not”); a hypothetical response option, in which the respondent volunteers an answer that falls on the response dimension but was not offered to the respondent (e.g., “pretty good”); an answer that gives a range (e.g., “good to very good”); and mitigating phrases that reduce the exactness, precision, or certainty of an answer and are offered as answers or parts of answers (e.g., “I guess excellent” or “just,” “maybe,” “about,” “put,” or “I’d say”).

Response latency, the time in seconds from the end of the interviewer's reading of the question until the respondent's first complete codable answer, has been used as an indicator of cognitive processing time, and in some cases longer times are associated with lower (or higher) data quality (Ehlen et al., 2007; Schaeffer and Dykema, 2004; Schaeffer et al., 2008; Yan and Tourangeau, 2008). In our analysis, response latency ends when the respondent provides one of the response options (excellent, very good, good, fair, or poor). The mean response latency was 1.73 seconds (standard deviation 2.42). We used the natural log of response latency in order to correct for the positive skew in the distribution of response latencies.

Interviewers often intervene in order to obtain a codable answer from a respondent following signs of hesitation or after the respondent provides an initial but inadequate answer, and such pre-emptive and follow-up behaviors by interviewers may indicate that a respondent is having difficulty answering (see Schaeffer and Maynard, 2002). For example, if the respondent provides a range (e.g., “good to very good”) as her initial answer, the interviewer might follow-up with the response options in order to obtain a single answer. Only 7% of our cases include interviewers speaking other than to administer the question or acknowledge the answer, so we created an index of other interviewer behaviors and then dichotomized

² The Kappas for the individual codes we used to construct our indices of interactional behaviors for this analysis are as follows: respondent token (0.969; there were 18 events which at least one coder coded as a token across the 30 cases double-coded for reliability analysis), respondent report/consideration (0.658; 7 events), respondent hypothetical response option, (0.496; 2 events), respondent range answer (0; 1 event), mitigating phrases such as “put,” or “I’d say” (0.963; 15 events), mitigating phrases such as “I guess excellent” or “just,” “maybe,” “about,” (0.855; 4 events), interviewer repeating response options (0.887; 5 events), and other interviewer follow-ups such as asking for a best guess or asking what health means to the respondent (1.0; 3 events). (The indices for nonparadigmatic sequence and exchanges are summaries of the interaction created by the authors rather than made up of behaviors coded by coders, and thus do not have any direct reliabilities associated with them. For budgetary reasons, response latencies were timed only when the original case was coded, thus we do not have any comparison between the original and double-coded case timings of response latency.)

³ Interviewers and interested researchers are not able to systematically observe uncontrollable aspects of the interview process when the survey is conducted over the telephone. As a result, it is possible that uncontrollable aspects of the interview process – such as the respondent's watching television, doing household chores or other activities, or having conversations with other household members – may affect some of the respondents' interactional measures in ways in which we cannot assess. We trained coders to code for instances when a person other than the respondent was talking or background noises were perceptible, and none of these situations were coded during the administration of the self-reported health item. Thus, we expect the effect of at least some of these uncontrollable aspects of the interview process to be minimal. We thank an anonymous reviewer for pointing this out.

the behaviors into a dummy variable (0 = no behavior other than reading the question from the interviewer, 1 = any other interviewer behavior).

Two variables in the analysis capture patterns in the interaction. Sequences are coded as paradigmatic when the interviewer's administration of the question is directly followed by the respondent's answer (with or without a preceding pause). The sequence may end there, or the interviewer may say "okay" or repeat the respondent's answer before the sequence ends (Schaeffer and Maynard, 1996). Any sequence that does not follow this pattern is nonparadigmatic (0 = paradigmatic, 1 = nonparadigmatic). The number of exchanges between the interviewer and respondent that occur in a question–answer sequence has also been found to be associated with lower data quality (Schaeffer and Dykema, 2004; see also van der Zouwen and Smit, 2004). We include a dummy variable for whether the interaction has more than one exchange or not (0 = no, 1 = yes).

4.2.2. Independent variables

The main independent variables are the respondent's cognitive ability, the respondent's answer to the self-reported health question, and a health inconsistency index. Cognitive ability is indicated by the respondent's IQ score (normalized) which was assessed during the respondent's junior year of high school (if available; if not, during the respondent's freshman year) using the Henmon-Nelson test of mental ability (mean 101.99, standard deviation 17.94) (Retherford and Sewell, 1988). Of the 355 respondents in this analysis, 102 reported "excellent" health, 125 reported "very good," 95 reported "good," 24 reported "fair," and 9 reported "poor." "Fair" and "poor" health ratings are combined into one category because of the small number of respondents in each category. Self-reported health is coded so that a higher value indicates worse health (e.g., "excellent" = 1 to "fair"/"poor" = 4).

Our health inconsistency index summarizes respondents' standing on indicators of functioning and presence of various health conditions. Functioning is assessed from a subscale of the Health Utilities Index (HUI3; <http://www.healthutilities.com>), focusing on the domains of the HUI3 that seemed the most relevant for functioning: ambulation, dexterity, and pain. Functioning scores range from .89 to 3, with a mean of 2.86 (standard deviation .31); scores are dichotomized in terms of mean functioning or less (22% of the sample) or greater than mean functioning (78% of the sample). Health conditions include arthritis, diabetes, high blood pressure, heart conditions (current or in past), cancer, stroke, and high blood sugar. Respondents reported having zero to five conditions, with a mean of 1.37 conditions (standard deviation 1.09). Because respondents are older, approximately 67 years at the time of the interview, and likely to have at least one condition, the measure of health conditions is dichotomized into zero or one condition (58% of the sample) versus two or more conditions. Using these two pieces of information, we code those with a low number of health conditions (zero or one) and high functioning (above the mean) as having consistent health information (50% of the sample); that is, their health status is probably clear, and it is probably relatively easy for them to select one of the response options indicating positive overall health – "excellent" or "very good." All other respondents – e.g., those with a high number of health conditions and high functioning, a low number of health conditions and low functioning, and a high number of health conditions and low functioning – have what we consider to be inconsistencies between health conditions and functioning, and it may be less clear to them which self-reported health response option most accurately reflects their status.⁴ The proportion of respondents classified as inconsistent varies as expected by answers to the self-reported health question: 29% of respondents answering "excellent," 43% of those answering "very good," 70% of those answering "good," and 85% of those answering "fair" or "poor."

5. Results

5.1. Interactional behaviors by self-reported health

First we investigate whether the interactional behaviors that occur during the self-reported health question–answer sequence vary by the answer given. Because the answer to the self-reported health question and the behaviors are co-produced, we expect that the behaviors that accompany answers to the self-reported health question will vary by the answer given. Specifically, we predict that there will be more problematic interactional behaviors when respondents report worse health.

Fig. 2 shows that in general, the proportion of respondents that produce problematic interactional behaviors is higher among respondents who report worse health. The exceptions are having any uncertainty indicators and a nonparadigmatic sequence, for which the proportions are lower for "fair/poor" compared to "good."

Fig. 3 plots the means and standard deviations of response latency (in seconds) for each self-reported health answer. The line down the center represents the mean response latency in seconds for each self-reported health answer, the gray band

⁴ It is also plausible that those with low functioning and a high number of health conditions have "consistent" health information, in that it is consistently not good. We also analyzed a trichotomous version of the health inconsistency index in the analyses reported below, with low functioning and a high number of health conditions as its own category. Results were similar for this category and the inconsistent category, and were thus combined in the analyses presented below. We think "consistently poor" combinations of functioning and conditions operate like inconsistency between functioning and conditions for a few reasons: it is easier to identify when health is good than when it is not; respondents with poor health and functioning face challenges in integrating this information to arrive at an overall judgment; in addition, because "excellent" is the first response option provided, respondents anchor to that first response option and tailor their response accordingly.

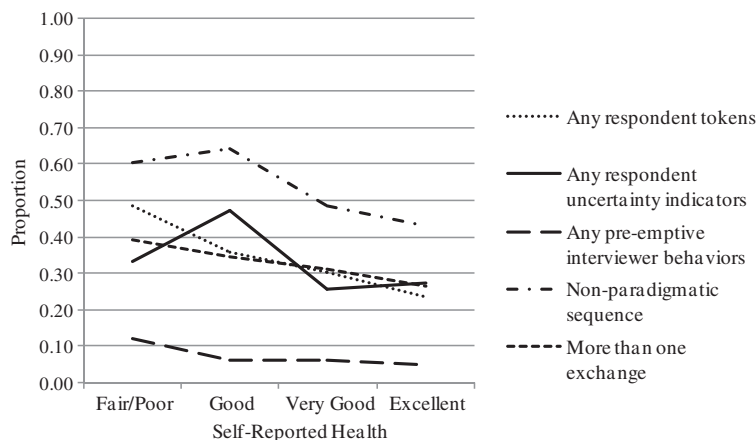


Fig. 2. Proportion of question–answer sequences showing each problematic interactional behavior by answer to the self-reported health question.

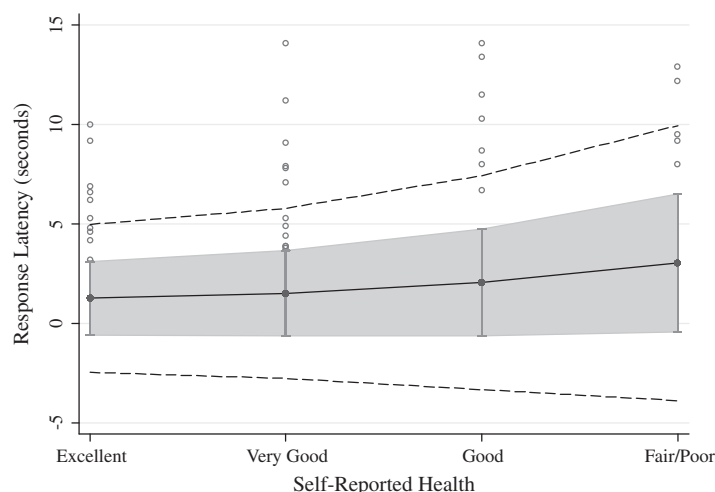


Fig. 3. Mean-standard deviation plot of response latency by self-reported health answer.

shows one standard deviation around the mean response latency for each answer, and the outer band of dashed lines shows two standard deviations around each mean. Each circle outside the second standard deviation range represents an individual case that falls beyond two standard deviations from the mean. Fig. 3 shows that the means and standard deviations of response latencies are larger for respondents with worse self-reported health. The pattern of means shows that the processing required to produce the summary requested by the self-reported health question requires more time for some of those in poorer health. The pattern of standard deviations suggests that the processing requirements for all respondents are heterogeneous and increasingly so as respondents report worse health, raising the possibility that respondents' health histories are increasingly heterogeneous with worse self-reported health.

We regress each of the interactional behaviors on self-reported health to investigate the first hypothesis, that the behaviors vary across the self-reported health response options. Because most of the behaviors are dichotomous variables, we use logistic regression and report the odds ratios. The exception is response latency, where we perform OLS regression and report the coefficient. We also use likelihood-ratio tests to determine whether self-reported health can be treated as an interval measure with evenly spaced categories or whether this leads to a loss of information about the association between the independent and dependent variables (Long and Freese, 2006).

Bivariate regression analyses demonstrate that the odds of each problematic interactional behavior significantly increase for respondents whose self-reported health is worse. The exception is any other interviewer behavior, which does not vary by respondents' self-reported health (see Table 3). Likelihood-ratio tests (not shown) demonstrate that the self-reported health response options can be treated as evenly spaced for all regressions except when the dependent variable is "any uncertainty indicators." Logistic regressions of any uncertainty indicators on dummy variables for the self-reported health response options show that the odds of producing any uncertainty behaviors are significantly lower when self-reported health is "excellent" or "very good" compared to "good," and the odds of producing any uncertainty behaviors are lower

Table 3
Regressions of behaviors on self-reported health.^a

	Mean or proportion	Coefficient or odds ratio	95% CI
<i>Respondent behaviors</i>			
Response latency (natural log of tenths of seconds)	−0.13 (1.20)	0.34***	(0.22, 0.47)
Any tokens [vs. none]	0.32	1.40**	(1.11, 1.78)
Any uncertainty indicators [vs. none] [“good” omitted]	0.33		
“excellent”		0.42**	(0.23, 0.76)
“very good”		0.40***	(0.23, 0.70)
“fair/poor”		0.56	(0.24, 1.27)
<i>Interviewer behaviors</i>			
Any other behaviors [vs. none]	0.07	1.30	(0.84, 2.01)
<i>Interactional behaviors</i>			
Nonparadigmatic sequence [vs. paradigmatic]	0.52	1.40**	(1.11, 1.75)
More than one exchange [vs. one exchange]	0.32	1.34*	(1.04, 1.73)

* $p < .05$.

** $p < .01$.

*** $p < .001$.

^a Mean and coefficient reported for response latency; proportion and odds ratios reported for all other interactional behaviors.

(but not significantly so) when self-reported health is “fair” or “poor” compared to “good” (see Table 3). In other words, self-reported health has a curvilinear relationship with any uncertainty behaviors, but for all of the other behaviors the relationship is linear, such that levels of the problematic behaviors are lower when better levels of health are reported and higher when worse levels of health are reported.

5.2. Predicting interactional behaviors during the self-reported health question–answer sequence from actual health

In the previous section we showed that there are some differences in the likelihood of the problematic interactional behaviors occurring for different levels of self-reported health. Next, we examine whether inconsistencies in respondents' health information predict behaviors during the self-reported health question–answer sequence. We regress each of the behaviors on the health inconsistency index, then perform the same regressions while controlling successively for the respondent's self-reported health, cognitive ability, and sociodemographic characteristics in order to determine whether the behaviors are associated with information about the complexity of respondent's health status beyond the self-reported health answer provided and other potential confounders.

When we predict behaviors produced during self-reported health question–answer sequence from the health inconsistency index, we find that each behavior is more likely to occur when the respondent has an inconsistent health history, and longer response latencies are positively associated with health inconsistencies (see Table 4). These relationships are slightly attenuated, but still statistically significant, when controlling for the respondent's answer to the self-reported health question. Controlling for cognitive ability does not lead to any changes in the sizes of the coefficients or odds ratios or their statistical significance (Table 4), nor does controlling for a range of 2004 sociodemographic characteristics including the respondent's gender, marital status, residence, educational attainment, and household income (not shown). Thus, interactional behaviors are associated with information about the respondent's health status – in this case, the complexity of the respondent's health status as indicated by inconsistencies in health conditions and functioning – beyond the answer to the self-reported health question, cognitive ability, and sociodemographic control variables.

5.3. Association between interactional behaviors and inconsistent health status within self-reported health answer

To determine whether interactional behaviors have predictive power beyond that of self-reported health, we examine the correlations between inconsistent health status and each interactional behavior among respondents who have the same answer to the self-reported health question. Table 5 shows that among respondents who reported “excellent” health, the behaviors were associated with health inconsistencies in the expected direction. For respondents who reported “very good” health, this was also the case with tokens, nonparadigmatic sequences, and exchanges. The significant correlations are evidence that even though all of these respondents answered “excellent,” for example, the interactional behaviors may provide additional information about the respondent's health status (in this case the complexity or inconsistency of health status) among these respondents. It is potentially important that these relationships appear most consistently in the “excellent” category. There is some limited experimental evidence that when “excellent” is the first response option offered, answers may be biased toward the positive end of the scale, and the relationship between self-reported health and number of health visits reduced (Means et al., 1989, pp. 18–19). If that were the case, information about the interaction could be important in correcting for this bias and improving construct validity.

Table 4Regression of behaviors on health inconsistency index.^a

	Bivariate		Controlling for SRH		Controlling for SRH, IQ	
	Coefficient or odds ratio	95% CI	Coefficient or odds ratio	95% CI	Coefficient or odds ratio	95% CI
<i>Respondent behaviors</i>						
Response latency (natural log of tenths of seconds)	0.47***	(0.23, 0.72)	0.27*	(0.01, 0.53)	0.27*	(0.01, 0.53)
Any tokens [vs. none]	1.94**	(1.23, 3.07)	1.65*	(1.01, 2.70)	1.66*	(1.02, 2.73)
Any uncertainty indicators [vs. none]	1.86**	(1.19, 2.93)	1.65*	(1.02, 2.68)	1.66*	(1.02, 2.70)
<i>Interviewer behaviors</i>						
Any other behaviors [vs. none]	3.89**	(1.41, 10.74)	3.79*	(1.31, 11.00)	3.78*	(1.31, 10.97)
<i>Interactional behaviors</i>						
Nonparadigmatic sequence [vs. paradigmatic]	2.34***	(1.53, 3.59)	2.05**	(1.30, 3.24)	2.09**	(1.32, 3.31)
More than one exchange [vs. one exchange]	2.01**	(1.22, 3.33)	1.77*	(1.03, 3.24)	1.77*	(1.03, 3.04)

* $p < .05$.** $p < .01$.*** $p < .001$.^a Coefficient reported for response latency; odds ratios reported for all other interactional behaviors.**Table 5**

Correlations of behaviors with health inconsistencies index by self-reported health.

	Excellent (<i>N</i> = 102)	Very good (<i>N</i> = 125)	Good (<i>N</i> = 95)	Fair/poor (<i>N</i> = 33)
<i>Respondent behaviors</i>				
Response latency (natural log of tenths of seconds)	0.25*	0.11	−0.01	−0.09
Any tokens [vs. none]	0.20*	0.20*	0.01	−0.27
Any uncertainty indicators [vs. none]	0.23*	0.06	0.08	−0.06
<i>Interviewer behaviors</i>				
Any other behaviors [vs. none]	0.25*	0.10	0.08	0.16
<i>Interactional behaviors</i>				
Nonparadigmatic sequence [vs. paradigmatic]	0.26**	0.25**	0.02	−0.17
More than one exchange [vs. one exchange]	0.21*	0.17†	−0.05	0.09

* $p < .05$.** $p < .01$.*** $p < .001$.† $p < .1$.

6. Discussion

Answers to the self-reported health question and the interactional behaviors we examined are co-produced. We expected and found more problematic interactional behaviors when respondents reported worse self-reported health. Furthermore, these behaviors appear to be related to a measure of health inconsistency, in that the interactional behaviors were more likely to occur when respondents had health inconsistencies, even after controlling for potential confounders. This finding indicates that the behaviors are associated with information about the respondent's health status – in this case, the complexity of the respondent's health status as indicated by inconsistencies between health conditions and functioning – beyond the respondent's answer to the self-reported health question, cognitive ability, and sociodemographic characteristics.

Additional analyses reveal that when the respondent's employment status is added to the regression of each interactional behavior on health inconsistency, self-reported health, cognitive ability, and sociodemographic characteristics, the effect size of the health inconsistency index diminishes for all interactional behaviors; the effect is no longer statistically significant for any except nonparadigmatic sequence and other interviewer behavior. Respondents who are currently employed are less likely to exhibit problematic interactional behaviors than respondents who are not currently employed, and we speculate that being currently employed might capture something about current health and cognitive status beyond that captured by our other measures. It is also possible that employed respondents are engaged more frequently in social interaction and as a result are afforded the opportunity to “practice” social interaction such that these respondents exhibit fewer problematic interactional behaviors, even when their health information is inconsistent. We find this to be a promising avenue of future research on interactional behaviors in the survey interview.

Finally, we also found that among respondents who reported “excellent” health, and to a lesser extent those who reported their health was “very good,” the behaviors were associated with inconsistencies between health conditions and functioning, indicating that interactional behaviors may be able to provide additional information about the respondent's health status among those with the same self-reported health.

Overall, we found that respondents whose health conditions and functioning were inconsistent were more likely to exhibit certain problematic interactional behaviors, providing evidence that these behaviors indicate difficulties in the response process (e.g., difficulties in combining information about disease and functioning or difficulties in mapping a resulting judgment onto one of the offered response options). It is plausible that the results of this analysis extend beyond the domain of health, giving us some idea of the behaviors that respondents exhibit when grappling with complex experiences, as well as the interactional patterns that may occur thereafter, such as interviewer follow-ups and nonparadigmatic sequences.

Furthermore, we find promising evidence that interactional behaviors might reveal important information about the respondent's health status. When limited information about health is collected in a survey, researchers using self-reported health as a measure of health might benefit from coding certain behaviors in the interaction as an indirect measure of health status or measurement error, particularly given that digital recordings of interviews are technically feasible in both telephone and face-to-face modes and are accepted by almost all respondents (typically more than 90%). Researchers might improve the predictive power of self-reported health by augmenting the answer to that question with measures of the interactional behaviors exhibited during the question–answer sequence. Future analyses should explore whether using these behaviors as additional control variables leads to improvements in models that use self-reported health as an independent or dependent variable.

Research has shown that respondents who answer a question with “don't know” or refuse are similar in their actual experience to respondents who provide an answer while expressing uncertainty, and that models to impute answers for those who “don't know” or refuse using information from those respondents who provide answers while expressing uncertainty perform better than models using information from all respondents (Mathiowetz, 1998). Thus, it is plausible that self-reported health answers could be more accurately imputed for item nonrespondents using data from respondents who exhibited certain interactional behaviors while answering the question.

Acknowledgments

An earlier version of this paper was presented at the 2010 annual meeting of the American Association for Public Opinion Research, Chicago, IL. The research reported here was supported in part by the National Institute of Child Health and Human Development (Center Grant R24 HD047873 and Training Grant T32 HD007014) and by the National Institute on Aging (Center Grant P30 AG017266 and the Wisconsin Longitudinal Study: Tracking the Life Course P01 AG021079). This research uses data from the Wisconsin Longitudinal Study (WLS) of the University of Wisconsin-Madison. Since 1991, the WLS has been supported principally by the National Institute on Aging (AG-9775 and AG-21079), with additional support from the Vilas Estate Trust, the National Science Foundation, the Spencer Foundation, and the Graduate School of the University of Wisconsin-Madison. A public use file of data from the Wisconsin Longitudinal Study is available from the Wisconsin Longitudinal Study, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, Wisconsin 53706 and at <http://www.ssc.wisc.edu/wlsresearch/data/>. All opinions and errors are the responsibility of the authors and do not necessarily reflect those of the funding agencies or helpful commentators. We are grateful to two anonymous reviewers for insightful comments and suggestions on earlier drafts.

References

- Benyamini, Yael, Leventhal, Elaine A., Leventhal, Howard, 1999. Self-assessments of health: what do people know that predicts their mortality? *Research on Aging* 21 (3), 477.
- Benyamini, Yael, Idler, Ellen L., Leventhal, Howard, Leventhal, Elaine A., 2000. Positive affect and function as influences on self-assessments of health expanding our view beyond illness and disability. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 55 (2), 107–116.
- Benyamini, Yael, Leventhal, Elaine A., Leventhal, Howard, 2000. Gender differences in processing information for making self-assessments of health. *Psychosomatic Medicine* 62, 354–364.
- Benyamini, Yael, Leventhal, Elaine A., Leventhal, Howard, 2003. Elderly people's ratings of the importance of health-related factors to their self-assessments of health. *Social Science & Medicine* 56 (8), 1661–1667.
- Bortfeld, Heather, Leon, Silvia D., Bloom, Jonathan E., Schober, Michael F., Brennan, Susan E., 2001. Disfluency rates in conversation: effects of age, relationship, topic, role, and gender. *Language and Speech* 44, 123–149.
- Canfield, Beth, Miller, Kristen, Beatty, Paul, Whitaker, Karen, Calvillo, Alfredo, Wilson, Barbara, 2003. Adult Questions on the Health Interview Survey – Results of Cognitive Testing Interviews Conducted April–May 2003. Hyattsville, MD: National Center for Health Statistics, Cognitive Methods Staff. <http://www.cdc.gov/QBANK/report%5CCanfield_NCHS_2003AdultNHISReport.pdf>.
- Cohen, Jacob, 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20 (1), 37–46.
- Draisma, Stasja, Dijkstra, Wil, 2004. Response latency and (para) linguistic expressions as indicators of response error. In: Presser, Stanley, Rothgeb, Jennifer M., Couper, Mick P., Lessler, Judith T., Martin, Elizabeth, Martin, Jean, Singer, Eleanor (Eds.), *Methods for Testing and Evaluating Survey Questionnaires*, Springer-Verlag, New York, pp. 131–148.
- Dykema, Jennifer, Lepkowski, James M., Blixt, Steven, 1997. The effect of interviewer and respondent behavior on data quality: analysis of interaction coding in a validation study. In: Lyberg, Lars, Biemer, Paul, Collins, Martin, de Leeuw, Edith, Schwarz, Norbert, Trewin, Dennis (Eds.), *Survey Measurement and Process Quality*. Wiley-Interscience, New York, pp. 287–310.
- Ehlen, Patrick, Schober, Michael F., Conrad, Frederick G., 2007. Modeling speech disfluency to predict conceptual misalignment in speech survey interfaces. *Discourse Processes* 44 (3), 245–265.
- Fleiss, Joseph L., 1981. *Statistical Methods for Rates and Proportions*. John Wiley, New York.
- Fowler Jr., Floyd J., Cannell, Charles F., 1996. Using behavioral coding to identify cognitive problems with survey questions. In: Schwarz, Norbert, Sudman, Seymour (Eds.), *Answering Questions: Methodology for Determining Cognitive, Communicative Processes in Survey Research*. Jossey-Bass, San Francisco, CA, pp. 15–36.

- Groves, Robert M., 1996. How do we know what we think they think is really what they think? In: Schwarz, Norbert, Sudman, Seymour (Eds.), *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. Jossey-Bass, San Francisco, CA, pp. 389–402.
- Groves, Robert M., Fultz, Nancy H., Martin, Elizabeth, 1992. Direct questioning about comprehension in a survey setting. In: Tanur Judith, M. (Ed.), *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*. Russell Sage Foundation, New York, pp. 49–61.
- Hauser, Robert M., 2005. Survey response in the long run: the Wisconsin Longitudinal Study. *Field Methods* 17 (1), 3–29.
- Hauser, Robert M., 2009. The Wisconsin Longitudinal Study: designing a study of the life course. In: Elder, G.H., Giele, J.Z. (Eds.), *The Craft of Life Course Research*. The Guilford Press, New York, pp. 29–50.
- Hess, Jennifer, Singer, Eleanor, Bushery, John M., 1999. Predicting test-retest reliability from behavior coding. *International Journal of Public Opinion Research* 11 (4), 346–360.
- Holbrook, Allyson, Cho, Young Ik, Johnson, Timothy, 2006. The impact of question and respondent characteristics on comprehension and mapping difficulties. *Public Opinion Quarterly* 70 (4), 565–595.
- Idler, Ellen L., Benyamini, Yael, 1997. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior* 38 (1), 21–37.
- Knäuper, Barbel, Belli, Robert F., Hill, Daniel H., Herzog, A.R., 1997. Question difficulty and respondents' cognitive ability: the effect on data quality. *Journal of Official Statistics* 13 (2), 181–199.
- Krause, Neal M., Jay, Gina M., 1994. What do global self-rated health items measure? *Medical Care* 32 (9), 930–942.
- Long, J. Scott, Freese, Jeremy, 2006. *Regression Models for Categorical Dependent Variables Using Stata*. Stata Press, College Station, TX.
- Mathiowetz, Nancy A., 1998. Respondent expressions of uncertainty: data source for imputation. *Public Opinion Quarterly* 62 (1), 47–56.
- Mathiowetz, Nancy A., 1999. Respondent uncertainty as indicator of response quality. *International Journal of Public Opinion Research* 11 (3), 289–296.
- Means, Barbara, Nigam, Arti, Zarrow, Marlene, Loftus, Elizabeth F., Donaldson, Molla S., 1989. *Autobiographical memory for health-related events*. Cognition and Survey Research, 6, 2, DHHS (PHS) 89-1077. Rockville, MD: National Center for Health Statistics.
- Retherford, Robert D., Sewell, William H., 1988. Intelligence and family size reconsidered. *Social Biology* 35 (1–2), 1–40.
- Schaeffer, Nora Cate, Dykema, Jennifer, 2004. A multiple-method approach to improving the clarity of closely related concepts: distinguishing legal and physical custody of children. In: Presser, Stanley, Rothgeb, Jennifer M., Couper, Mick P., Lessler, Judith T., Martin, Elizabeth, Martin, Jean, Singer, Eleanor (Eds.), *Methods for Testing and Evaluating Survey Questionnaires*, Springer-Verlag, New York, pp. 475–502.
- Schaeffer, Nora Cate, Dykema, Jennifer, 2011. Coding the behavior of interviewers and respondents to evaluate survey questions: a response. In: Madans, Jennifer, Miller, Kristen, Willis, Gordon, Maitland, Aaron (Eds.), *Question Evaluation Methods: Contributing to the Science of Data Quality*. Wiley and Sons, New York, pp. 23–38.
- Schaeffer, Nora Cate, Maynard, Douglas W., 1996. From paradigm to prototype and back again: interactive aspects of cognitive processing in survey interviews. In: Schwarz, Norbert, Sudman, Seymour (Eds.), *Answering Questions: Methodology for Determining Cognitive, Communicative Processes in Survey Research*. Jossey-Bass, San Francisco, pp. 65–88.
- Schaeffer, Nora Cate, Maynard, Douglas W., 2002. Occasions for intervention: interactional resources for comprehension in standardized survey interviews. In: Maynard, Douglas W., Houtkoop-Steenstra, Hanneke, Schaeffer, Nora C., van der Zouwen, Johannes (Eds.), *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, Wiley, New York, pp. 261–280.
- Schaeffer, Nora Cate, Dykema, Jennifer, Garbarski, Dana, Maynard, Douglas W., 2008. Verbal and paralinguistic behaviors in cognitive assessments in a survey interview. In: *JSM Proceedings, Survey Research Methods Section*. American Statistical Association, Alexandria, VA, pp. 4344–4351.
- Schegloff, Emanuel A., 1981. Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences. In: Tannen, D. (Ed.), *Analyzing Discourse* (Georgetown University Roundtable on Languages and Linguistics 1981), Georgetown University Press, Washington, DC, pp. 71–93.
- Sewell, William H., Hauser, Robert M., Springer, Kristen W., Hauser, Taissa S., 2004. As we age: the Wisconsin longitudinal study, 1957–2001. In: Leicht, K. (Ed.), *Research in Social Stratification and Mobility*, vol. 20. Elsevier, London, pp. 3–111.
- Tourangeau, Roger, Rips, Lance J., Rasinski, Kenneth, 2000. *The Psychology of Survey Response*. Cambridge University Press, UK.
- van der Zouwen, Johannes, Smit, Johannes H., 2004. Evaluating survey questions by analyzing patterns of behavior codes and question–answer sequences: a diagnostic approach. In: Presser, Stanley, Rothgeb, Jennifer M., Couper, Mick P., Lessler, Judith T., Martin, Elizabeth, Martin, Jean, Singer, Eleanor (Eds.), *Methods for Testing and Evaluating Survey Questionnaires*, Springer-Verlag, New York, pp. 109–130.
- Wisconsin Longitudinal Study (WLS) [graduates]: 2004 Version 12.10. [machine-readable data file] / Hauser, Robert M; Hauser, Taissa S. [principal investigator(s)]. Madison, WI: University of Wisconsin-Madison, WLS. [distributor]; <<http://www.ssc.wisc.edu/wlsresearch/documentation/>>.
- Yan, Ting, Tourangeau, Roger, 2008. Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology* 22, 51–68.