Verbal and Paralinguistic Behaviors in Cognitive Assessments in a Survey Interview

Nora Cate Schaeffer[1], Jennifer Dykema[2], Dana Garbarski[3], Douglas W. Maynard[4]
[1]Department of Sociology, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706
[2]University of Wisconsin Survey Center, 1800 University Avenue, Madison, WI 53726
[3]Department of Sociology, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706
[4]Department of Sociology, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706

Abstract
Survey researchers have examined the interaction between interviewers and respondents for behavioral evidence of difficulties with the survey task or measurement error or both. Using detailed coding of interaction during cognitive assessments conducted by telephone in the Wisconsin Longitudinal Study, we examine the relationships between behaviors by the respondent and performance on a letter fluency task designed to measure verbal semantic memory and a digit ordering task designed to measure fluid intelligence and working memory span. We find that longer response latencies are associated with poorer performance on the letter fluency task but with better performance on the digit ordering task. "Don't know" answers are unrelated to success in the letter fluency task but are associated with poorer performance on the digit ordering task. Results for other behaviors are more complex and less conclusive.

Key Words: behavior coding, cognitive assessment, data quality, disfluencies of speech, interaction coding, interviewer-respondent interaction, measurement error, paralinguistic measures, response accuracy, survey interviewing

1. Introduction

Standardized survey interviewing is designed to produce a paradigmatic interaction: The interviewer asks a question, the respondent answers, the interviewer acknowledges the answer, and the two participants move to the next question (Schaeffer and Maynard, 1996). When the behavior of the participants deviates from this paradigm, it is likely to be due to the difficulty of the task or to the ability of the respondent, with the result that some behaviors are associated with reduced data quality. Survey methodologists who study interaction in the survey interview have drawn on the work of conversation analysts, psychologists, and others to identify behaviors and other features of interaction in standardized interviews that are associated with measurement error or other aspects of data quality. For example, various paralinguistic (nonverbal) and verbal behaviors associated with decreased accuracy in record-check studies include the following: an increased number of exchanges between the interviewer and respondent, follow-up or probing by the interviewer, and laughter (Schaeffer & Dykema, 2004); the respondent interrupting or seeking clarification (Dykema, Lepkowski, and Blixt 1997); the number of words the respondent uses (Draisma & Dijkstra, 2004); the use of words that mitigate an answer by expressing lack of knowledge, doubt, uncertainty, approximation, or qualification (Mathiowetz, 1999; Dykema, Blixt & Lepkowski, 1997; Schaeffer & Dykema, 2004; Draisma & Dijkstra, 2004); and response latency, that is, the time between the end of a question and the beginning of the answer (Schaeffer & Dykema, 2004; Draisma & Dijkstra, 2004; see also Ehlen, Schober & Conrad, 2007). Many of these behaviors appear as disfluent talk, that is, talk that is broken or mitigating, hesitant, or involves repairs or restarts. Some of these behaviors have been used to identify difficult questions (e.g., Fowler, 1992), but the likelihood of "don't know" answers has also been found to be associated with the respondent's cognitive ability; furthermore the difficulty of the question interacts with the ability of the respondent, so that respondents with lower cognitive ability are more likely than other respondents to answer "don't know" to difficult questions (Knauper, Belli, Hill, & Herzog, 1997).

This previous research suggests that paralinguistic behaviors and other behaviors that accomplish similar actions (such as words that mitigate answers) sometimes provide information about the respondent's cognitive processing or cognitive ability and, therefore, about a source of measurement error. A more precise understanding of when these behaviors provide such information requires systematic comparisons among candidate behaviors in an analysis that uses a criterion. Our analysis is embedded in a project that examines interaction during telephone interviews of older adults that included assessments of cognitive ability. Thus, we are able to examine the behaviors produced during different cognitive tasks. In addition, scores on cognitive assessments provide a criterion, so that we can describe whether the behaviors are associated differently with different cognitive abilities. A general hypothesis suggested by this literature would be that some paralinguistic and verbal behaviors indicate cognitive troubles that might be caused by a difficult task, limited ability, or an interaction between the two. Because we are observing behavior during cognitive assessments, the distinction between troubles caused by the difficulty of the task and by the respondent's ability is blurred: To the extent that the cognitive assessment succeeds in its goal of challenging a specific cognitive ability, when a respondent encounters difficulty in accomplishing the task, it is, presumably, because of his or her lower cognitive ability. Thus, overall, we expect more of these behaviors to be produced in an interaction that also produces lower cognitive scores. However, to the extent that the cognitive tasks tap different cognitive abilities, a particular behavior could be associated with success on one task but not on another. Furthermore, the tasks may, in the course of their performance, evoke behaviors that are associated with the exercise of a cognitive ability or the interactional structuring of its performance, but not with the level of cognitive ability. Thus, because verbal production is the cognitive task in the letter fluency task, we might expect utterances – both fluent and disfluent -- to be produced at a higher rate during that task than during digit ordering. In addition, some behaviors – such as "uh" or "okay" – accomplish many different social actions, and this fact could attenuate observed relationships with cognitive ability.

2. Data and Coding System

2.1 Wisconsin Longitudinal Study

We analyze a subset of cases from the Wisconsin Longitudinal Study (WLS), a longitudinal study of graduates from Wisconsin high schools in 1957. In the 2004 follow-up telephone interviews all respondents were asked for permission to record the interview, and almost all agreed. The WLS sample is divided into independent replicates. We randomly selected 138 interviewers from among those who conducted interviews in the even-numbered replicates to ensure that our cases would span the field period. Within interviewer, we identified respondents for whom records indicated that the interview had been recorded and sorted respondents by IQ measured in high school. For each interviewer we attempted to select five cases: the two respondents with the highest IQ, the two with the lowest, and the respondent with the median IQ. The order of cases was randomized before the cases were assigned to coders; for funding reasons, only the first 370 cases completed by 79 interviewers were coded; all but 10 interviewers in the final sample are represented by four or five respondents. For the nine respondents whose cognitive assessments or recordings were missing or uncodable, despite the notation in the records, we selected another respondent interviewed by the same interviewer in the five cases where that was possible; a handful of cases in the sample refused the cognitive tasks.

We coded interaction during two cognitive assessments that tap very different cognitive abilities (Lachman and Tun 2008). In the "letter fluency" task, which measures verbal semantic memory, respondents were given one minute "to say as quickly as you can all of the words you can think of that begin with" the letter "l" or "f," with the letter being randomly assigned (Borkowski, Benton, and Spreen 1967; Monsch, Bondi, Butters, Paulsen, Salmon, Brugger, and Swenson 1994). The "digit ordering" task is a modification of the (WAIS-III) digit backward subtest, which is thought to measure fluid intelligence and working memory span (Wechsler, 1997; Yonker, Hauser, Freese, 2007). In the interview respondents were presented with sets of unordered digits and asked to state them in ascending order. The task contained six levels, with each subsequent level including an additional digit. At the first level, respondents were presented with a set containing three unordered digits. A respondent who ordered the digits correctly at a given level was then presented with the next larger set of digits, up to the sixth and final level, which contained a set of eight digits. A respondent who ordered the digits incorrectly was given a second chance with a new string of the size they failed to order. The task ended if the respondent incorrectly ordered digits during a second chance or if the respondent correctly ordered all eight digits from the final set. The analyses for the letter fluency task are based on 353 cases; those for the digit ordering task use 352 cases.

2.2 Interaction Coding System and Independent Variables

To develop our coding system, we began with a conversation analysis of a sample of 50 transcripts (Schaeffer & Maynard, 2008; Gathman, Maynard, & Schaeffer, 2008) to identify events for coding (Dykema et al. 2007). Coding was done from transcripts using the Sequence Viewer program (Dijkstra 2002). We identified over 100 behaviors for coding based on our analysis of the transcripts, detailed examination of the interviews used for the conversation analysis, and the literatures on interaction in survey interviews. We coded these events at a very fine level of detail because prior research did not indicate which behaviors might be the most common, which might be characteristic of different cognitive tasks, or which might be used in similar ways by actors. Only by coding detailed behaviors could we discover which behaviors were common and which were not and which behaviors were similarly related to criteria.

In this paper we report about the following behaviors: response latency (the time in tenths of seconds between the end of the stimulus item and the first answer[1]); repairs (e.g., the respondent restarts a word); thinking phrases (e.g., "let's see"); mitigators; tokens and particles; and affirmations. Mitigators include words or phrases that express distance (e.g., "I'd say…"), uncertainty (e.g., "I guess"), not knowing (e.g., "don't know"), and approximation (e.g., "about"). Tokens include "well," "uh," "oh," "er," and a sizeable group of other tokens that did not fall into these groups. (We document the presence of this latter group here, but exploring its contents and meaning require later analysis.) Affirmations included "yes," "okay," "uh-huh," and "right." Coders were trained with detailed descriptions for each class of event. Figure 1 shows part of the coding manual for the event "token;" detailed additional instructions and examples explained what other tokens should be included in each group. The list of all behaviors we analyze here, which represent only a small portion of the coding system, appears in Tables 1 and 2. Although the rationale for including the behavior in our analysis is straightforward in most cases, our treatment of the affirmations perhaps requires some discussion. Most of the behaviors examined here – particularly the tokens and mitigators – occur in varied environments and can be used to accomplish many different actions. This is also true of the affirmative utterances. For example, after the standardized questions in the health section of the interview (which we do not analyze here), we commonly observe a pair of "okays," one acknowledging an answer, the other announcing the beginning of the next health question. In contrast, in the letter fluency task, a respondent might use "okay" to launch an initial string of words: "okay lake like um leech uh lint ah lintall," and in both tasks respondents use "okay" to acknowledge comments or return to the task after a brief exchange with the interviewer:

Interviewer: You got a little bit time left.
Respondent: okay.

For the letter fluency task, our analysis uses a dummy variable that indicates whether the respondent exhibited the behavior at all during the task, which lasted approximately one minute. For the digit ordering task, we use a dummy variable that indicates whether the behavior occurred on the penultimate item, that is, the item before the last item administered to the respondent.

2.3 Dependent Variables and Analytic Models[2]

For letter fluency we use as dependent variables the number of correct words (range = 0 to 26, mean = 11.24, median = 11) and the number of incorrect words (range = 0 to 14, mean = 1.96, median = 1) separately. In the case of the digit ordering task, approximately one-fifth (19%) of the respondents in this study completed it successfully. For the purposes of analysis the digit ordering task is scored in the following manner: For each level completed successfully on the first try, the respondent receives ten points for every digit ordered correctly. For example, a respondent who completed the first level correctly during the initial attempt is awarded 30 points. A respondent who answered incorrectly during the first attempt but succeeds during the second chance is awarded half of the full points for the level (e.g., 15 points for successfully completing level one on a second try). Points are accumulated as respondents progress from level to level and the total ranges in value from 0 points (for failing to order any sets of digits correctly; 7%) to 330 points (for ordering all six sets correctly during the initial attempt; 7%) (mean = 171.01, s. d. = 89.84, median = 162.5).

We describe the data by presenting simple bivariate regressions with the dummy variables for the behaviors as independent variables. We use dummy variables because the incidence of the behaviors varies within and between tasks, and because the opportunity for exhibiting the behaviors varied between the tasks. In Models 1 and 2 we predict the number of correct and incorrect words in the letter fluency task from the behaviors in the letter fluency task. We

[1] In the letter fluency task, response latencies are timed from the end of the interviewer saying "start now" until the *beginning* of the respondent's first answer that could be scored (word beginning with "F" or "L"). In the digit ordering task, response latencies are timed from the end of the interviewer's reading of the digit ordering until the *beginning* of the first digit the respondent gives as their reordering of the digits. We also timed response latencies from the end of the respective stimuli to the *end* of the first answer given in each task. These are correlated with the response latencies we used in the final analysis .94 in letter fluency and .997 in digit ordering.

[2] We indicate significance levels of tests as follows: * = $p < .10$, ** = $p < .05$, *** = $p < .01$

examine the number of incorrect words, because it is not obvious whether such words are substitutes for other sorts of verbal behaviors or indicate a lower level of fluency. In Model 3 we predict the digit ordering score from behaviors on the penultimate item, that is, the last item successfully completed. In Model 4, the digit ordering score is predicted from behaviors in the letter fluency task, which took place before the digit ordering task; this model allows us to examine whether behaviors during one type of cognitive demand (letter fluency) predict success under a different sort of cognitive demand (digit ordering). In the first three models, the behaviors and criterion are produced by the actors concurrently, and the coefficients simply summarize the relationship. The final model addresses whether the behaviors in one cognitive assessment predict success in a subsequent, different, cognitive task. IQ measured in high school is positively and significantly correlated with the number of correct words (0.33**) and the score for digit ordering (0.33**); the number of correct words is positively correlated with the score for digit ordering (0.24**). The number of incorrect words has a modest negative relationship with IQ (-0.10*) but is essentially uncorrelated with the number of correct words (-0.07) and the score for digit ordering (-0.06).

3. Results

The first columns in Tables 1 and 2 show the proportion of respondents exhibiting the behaviors at least once – in the letter fluency task, during all the digit ordering items the respondent completed, and during the penultimate digit ordering item.[3] As noted earlier, it is difficult to compare the incidence of behaviors across cognitive assessments: the length of time was held approximately constant for the letter fluency task, but respondents varied in how much of that time they filled with producing words and how much of the time they were silent; in the digit ordering task, the number

Table 1: Bivariate Regressions of Scores from Letter Fluency Task on Behaviors of Respondents

		Number of Correct Words			Number of Incorrect Words		
Behavior	Proportion	Coefficient	s.e.		Coefficient	s.e.	
Response latency	NA	-0.24	0.09	***	0.13	0.05	***
Repair	0.16	-0.46	0.64		0.24	0.34	
Thinking phrase	0.20	-0.52	0.60		1.01	0.32	***
Mitigators							
Distancing	0.06	0.45	1.00		0.20	0.54	
Uncertainty	0.12	-0.36	0.74		1.21	0.39	***
Don't know	0.09	-1.36	0.84		0.40	0.45	
Approximation	0.02	-4.48	1.69	***	-0.39	0.91	
Tokens							
Well	0.04	-1.37	1.26		0.20	0.67	
Uh	0.87	0.23	0.70		0.83	0.37	**
Oh	0.37	-1.33	0.49	***	0.55	0.26	**
Er	0.01	NA			NA		
Other token	0.21	-0.86	0.58		0.50	0.31	
Affirmations							
Yes	0.06	-1.85	1.02	*	-0.06	0.55	
Okay	0.24	-0.44	0.56		0.60	0.30	**
Uh-huh	0.02	0.33	1.70		-0.69	0.91	
Right	0.00	NA			NA		
Other affirmation	0.00	NA			NA		

Note: Response latency is measured in tenths of seconds. Other independent variables are coded 0, 1.The mean response latency for the letter fluency task is 1.98 (s.d. = 2.63).

of items (that is, the number of strings of digits) varied across respondents, and so the opportunity for exhibiting the behavior varied across respondents. Although the structure of the data (as just summarized) makes formal hypothesis

[3] We omit regression results if 1 percent or less of the sample did not exhibit the behavior.

testing difficult, when we compare the incidence of behaviors during the letter fluency task with their incidence during digit ordering, it appears plausible that thinking phrases and the "uh" token might be relatively more common during the letter fluency task and that uncertainty and "don't know" mitigators might be relatively more common when respondents are ordering digits.

Tables 1 and 2 also present the results from a series of bivariate regressions of the cognitive scores on the behaviors of respondents. Table 1 shows results for regressing the number of correct and incorrect words on behaviors exhibited during the letter fluency task; Table 2 shows results for regressing digit ordering scores on behaviors exhibited during the penultimate item and then on behaviors exhibited during the letter fluency task. If the behaviors were general indicators of cognitive troubles associated with all cognitive abilities, the same behaviors that predict success in letter fluency would predict success in digit ordering. However, this expectation is disconfirmed, at least for response latency. An overall impression one might form when listening to the recordings of these two tasks is that respondents are relatively noisy during the letter fluency task, as though keeping a channel for verbalizing associations open, and relatively quiet while reordering the digits, as though talking might disturb the contents of working memory. The different role of silence in the two tasks is suggested by the sign of the coefficient for the response latency: A longer response latency is associated with a lower number of correct words and a higher number of incorrect words in the letter fluency task (Table 1), but with a higher score when ordering digits (Table 2). Although discussions in the literature have considered whether a longer latency indicates cognitive problems or deeper thought, these measurements taken during different cognitive tasks demonstrate that response latency can indicate either – depending on the cognitive ability being recruited. The distinctiveness of these two response latencies is further confirmed by the negative relationship between the response latency from the letter fluency task and the digit ordering score (Table 2): The relationships suggest that some respondents who struggle with letter fluency also struggle with digit ordering.

Table 2: Bivariate Regressions of Scores from Digit Ordering Task on Behaviors of Respondents

Behavior	Proportion Through Penultimate Item	Proportion On Penultimate Item	On Behaviors During Penultimate Item			On Behaviors During Letter Fluency		
			Coefficient	s.e.		Coefficient	s.e.	
Response latency	NA	NA	5.11	1.97	***	-3.61	1.81	**
Repair	0.14	0.08	-12.18	17.43		3.60	12.95	
Thinking phrase	0.04	0.02	19.43	32.16		23.99	12.02	**
Mitigators								
Distancing	0.09	0.03	-13.55	26.42		31.48	20.53	
Uncertainty	0.24	0.12	6.45	14.94		20.24	14.99	
Don't Know	0.22	0.15	-33.57	13.18	**	-14.32	16.81	
Approximation	0.01	0.01	NA			-20.34	34.14	
Tokens								
Well	0.03	0.01	NA			56.17	25.11	**
Uh	0.42	0.25	25.17	10.99	**	16.67	14.13	
Oh	0.31	0.15	24.99	13.45	*	-11.95	9.88	
Er	0.04	0.01	NA			NA		
Other token	0.13	0.07	-13.79	18.65		9.89	11.82	
Affirmations								
Yes	0.09	0.03	20.09	26.40		14.78	21.09	
Okay	0.18	0.09	18.70	16.89		-6.84	11.24	
Uh-huh	0.02	0.01	NA			-40.02	34.09	
Right	0.02	0.00	NA			NA		
Other affirmation	0.003	0.00	NA			NA		

Note: Response latency is measured in tenths of seconds. Other independent variables are coded 0, 1.The mean response latency for the digit ordering task is 2.81 (s. d. = 2.41)

The findings are more complicated for the other behaviors. The occurrence of repairs does not predict success in either cognitive task. The presence of a thinking phrase during letter fluency or on the penultimate digit ordering item does not predict success on that task, but does predict the number of incorrect words. Three of the four behaviors we describe as mitigators have negative relationships with the score on letter fluency, although only for approximation – the rarest of the four – is the relationship significant. Similarly two of the three mitigators have negative relationships with the score for digit ordering, and the relationship is significant for “don’t know.” In both of these cases the presence of mitigators is associated with a lower score, though the relationship is only large enough to achieve significance in two instances. Four of the five types of tokens produced during the letter fluency task fail to predict the number of correct words during that task. The exception is “oh,” which has a significant negative relationship with the number of correct words. We coded “oh” separately because it is sometimes deployed as a “change-of-state token” (Heritage 1984). We have examples of “oh” introducing comments about the difficulty of the letter fluency task (“ooh, it’s a hard one”), which are compatible with this negative relationship. In contrast, the tokens “uh” and “oh” have almost identical positive and significant relationships with the respondent’s score during digit ordering. In the digit ordering task, these tokens might be used interchangeably or in complementary ways by respondents; for example, “uh” may be used in the same way as thinking phrases, and “oh” could be used to recognize an error that is then corrected: “zero three oh zero one three.” The affirmative words have no substantial associations with either the number of correct words or the digit ordering score.

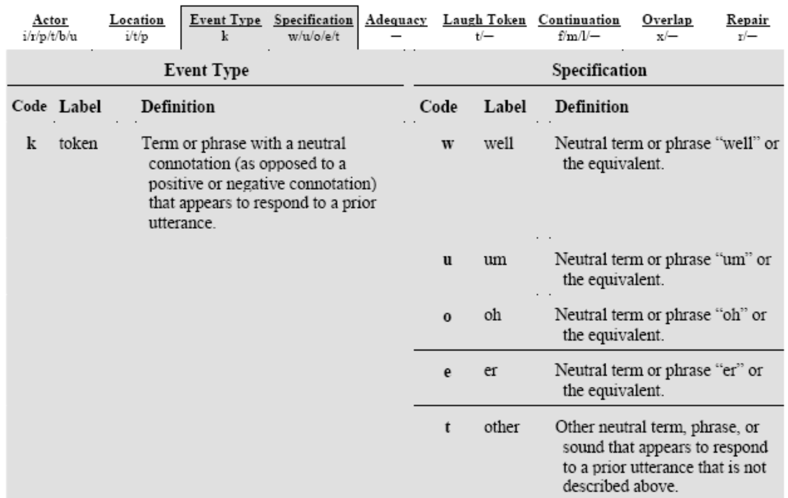
During the letter fluency task, several behaviors – thinking phrases, uncertainty mitigators, the tokens “uh” and “oh,” and the affirmative “okay” – predict the number of incorrect words, which is essentially uncorrelated with the number of correct words. The positive relationship between the number of incorrect words and the token “oh” is generated by comments like this one: “franken oh that’s a name Frankenstenein…fair oh I said that already.” It also seems plausible that for some respondents the words scored as incorrect serve a cognitive and interactional function similar to that served by these other behaviors.

Two behaviors exhibited during the letter fluency task – the use of thinking phrases and of “well” – predict the score in digit ordering. As was the case with response latency, these behaviors have negative (though non-significant) relationships with the score during the letter fluency task. The pattern of results supports the interactional distinctiveness of these cognitive abilities.

4. Discussion

Survey researchers have used behaviors of the type examined here to identify survey questions that cause problems for respondents. Such problems could originate in technical properties of the question (e.g., the wording or construction of the item), classification judgments (e.g., do reading glasses count as “glasses”), in the fit between complex concepts in the question and the respondent’s situation or biography (e.g., the question does not make clear whether someone on paid maternity leave should report having a job), in the challenge presented by the task (e.g., reporting all medical visits in the previous year), in the respondent’s ability (e.g., memory), in the respondent’s motivation, or in some interaction among these factors. Thus, whether problems occur or not will depend partly on characteristics of the population; for example, because “0” is an easy number to remember, if no respondent has medical visits to report, even a vague or complicated question about medical visits is unlikely to generate evidence of problems. In addition, whether such problems give rise to observable traces in the behavior of participants will depend partly on the social structuring of the task.

This first look at these new data suggests that understanding behaviors of the sort examined here may require making fine distinctions. The patterning in our results suggests that if further analysis can identify which mitigators or tokens respondents use interchangeably, stronger results would be obtained by combining them. Such decisions require attention to both conversation analytic issues (e.g., the use of “well” to maintain topical coherence, Schiffrin, 1985) as well as statistical issues (e.g., the way the use of “well” is associated with other behaviors and with potential criteria). The different implication of response latency in the two cognitive tasks we examine shows that the meaning of response latency and its relationship to success may depend on the task in which it occurs. Similarly, a response latency that is either relatively short or relatively long may be associated with errors in a response task that requires the person to make classification judgments as part of a response (e.g., does a lamp count as “furniture”) (Ehlen, Schober, and Conrad, 2007). There does not seem to be a simple relationship between the production of the other behaviors we examine and either of our measures of cognitive ability. Similarly, disfluent behaviors such as these “were not highly significant predictors” of response errors due to classification failures in a laboratory setting (Ehlen et al. 2007, p. 263). The meaning of such behaviors is likely to be specific to the type of task within which the behavior occurs: As we find here, a “don’t know” answer may be a meaningful indicator of error in a task that taxes memory, but not in a task that draws on other abilities. Similarly, a pause may mean something different in a task that requires associative thinking, applying semantic knowledge to classify experiences, or searching autobiographical memory. Many survey tasks draw on all these abilities. In addition, the results for the tokens "oh" and "uh" suggests that the interactional performance of the task may influence how such tokens are deployed, and, hence, their meaning.



Actor	Location	Event Type	Specification	Adequacy	Laugh Token	Continuation	Overlap	Repair
i/r/p/t/b/u	i/t/p	k	w/u/o/e/t	—	t/—	f/m/l/—	x/—	r/—

Event Type			Specification		
Code	**Label**	**Definition**	**Code**	**Label**	**Definition**
k	token	Term or phrase with a neutral connotation (as opposed to a positive or negative connotation) that appears to respond to a prior utterance.	**w**	well	Neutral term or phrase "well" or the equivalent.
			u	um	Neutral term or phrase "um" or the equivalent.
			o	oh	Neutral term or phrase "oh" or the equivalent.
			e	er	Neutral term or phrase "er" or the equivalent.
			t	other	Other neutral term, phrase, or sound that appears to respond to a prior utterance that is not described above.

Figure 1: Portion of page from coding manual for the event token

Acknowledgments

The research reported here was supported in part by the National Institute of Child Health and Human Development, (Center Grant R24 HD047873), by the National Institute on Aging (Center Grant P30 AG017266 and the Wisconsin Longitudinal Study: Tracking the Life Course P01 AG021079), by the Wisconsin Center for Demography and Ecology, by the Wisconsin Center for Demography of Health and Aging, by grants from the Graduate School Research Committee to Schaeffer and Maynard, and by the University of Wisconsin Survey Center (UWSC). Nora Cate Schaeffer and Douglas W. Maynard are co-investigators on the WLS Supplement "Cognition and Interaction in Surveys of Older Adults." We thank Wil Dijkstra and Yfke Ongena for patiently answering our questions about Sequence Viewer and Wil Dijkstra for generously making modifications to Sequence Viewer (http://www.sequenceviewer.nl/).

References

Borkowski, J. G., A. L. Benton, and O. Spreen. 1967. "Word Fluency and Brain Damage." *Neuropsychologica* 5:135-40.

Dijkstra, Wil. 2002. "Transcribing, Coding, and Analyzing Verbal Interactions in Survey Interviews." Pp. 401-26 in *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, edited by Douglas W. Maynard, Hanneke Houtkoop-Steenstra, Johannes van der Zouwen, and Nora Cate Schaeffer. New York: Wiley.

Draisma, Stasja and Wil Dijkstra. 2004. "Response Latency and (Para)Linguistic Expression As Indicators of Response Error." Pp. 131-48 in *Methods for Testing and Evaluating Survey Questionnaires*, edited by Stanley Presser, Jennifer M. Rothgeb, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer. New York: Springer-Verlag.

Dykema, Jennifer, Dana Garbarski, Nora Cate Schaeffer, and Douglas W. Maynard. 2007. "Code Manual for Codifying Interviewer-Respondent Interaction in Surveys of Older Adults." University of Wisconsin-Madison.

Dykema, Jennifer, James M. Lepkowski, and Steven Blixt. 1997. "The Effect of Interviewer and Respondent Behavior on Data Quality: Analysis of Interaction Coding in a Validation Study." Pp. 287-310 in *Survey Measurement and Process Quality*, edited by Lars Lyberg, Paul Biemer, Martin Collins, Edith de Leeuw, Cathryn. Dippo, Norbert Schwarz, and Dennis Trewin. N.Y.: Wiley-Interscience.

Ehlen, Patrick, Michael F. Schober, and Frederick G. Conrad. 2007. "Modeling Speech Disfluency to Predict Conceptual Misalignment in Speech Survey Interfaces." *Discourse Processes,* 44(3):245-65.

Fowler, Floyd J. Jr. 1992. "How Unclear Terms Affect Survey Data." *Public Opinion Quarterly* 56(2):218-31.

Gathman, E. H. C., Douglas W. Maynard, and Nora Cate Schaeffer. 2008. "The Respondents Are All Above Average: Compliment Sequences in a Survey Interview." *Research on Language and Social Interaction* 41(3):271-301.

Heritage, John. 1984. "A Change of State Token and Aspects of Its Sequential Placement." Pp. 299-345 in *Structures of Social Action: Studies in Conversation Analysis*, edited by J. M. Atkinson and John Heritage. Cambridge, MA: Cambridge University Press.

Knauper, Barbel, Robert F. Belli, Daniel H. Hill, and A. R. Herzog. 1997. "Question Difficulty and Respondents' Cognitive Ability: The Effect on Data Quality." *Journal of Official Statistics* 13(2):181-99.

Lachman, Margie E. and Patricia A. Tun. 2008. "Cognitive Testing in Large-Scale Surveys: Assessment by Telephone." Pp. 5006-522 in *Handbook on Cognitive Aging: Interdisciplinary Perspectives*, edited by S. Hofer and Duane Alwin. Thousand Oaks: Sage.

Mathiowetz, Nancy A. 1999. "Expressions of Respondent Uncertainty As Indicators of Data Quality." *International Journal of Public Opinion Research* 11(3):289-96.

Monsch, A. U., M. W. Bondi, N. Butters, J. S. Paulsen, D. P. Salmon, P. Brugger, and M. R. Swenson. 1994. "“A Comparison of Category and Letter Fluency in Alzheimer’s Disease and Huntington’s Disease.”." *Neuropsychology* 8:25-30.

Schaeffer, Nora Cate and Jennifer Dykema. 2004. "A Multiple-Method Approach to Improving the Clarity of Closely Related Concepts: Distinguishing Legal and Physical Custody of Children." Pp. 475-502 in *Methods for Testing and Evaluating Survey Questionnaires*, edited by Stanley Presser , Jennifer M. Rothgeb, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer. New York: Springer-Verlag.

Schaeffer, Nora Cate and Douglas W. Maynard. 2008. "The Contemporary Standardized Survey Interview for Social Research." Pp. 31-57 in *Envisioning the Survey Interview of the Future*, edited by Frederick G. Conrad and Michael F. Schober. Hoboken, NJ: Wiley.

Schaeffer, Nora Cate and Douglas W. Maynard. 1996. "From Paradigm to Prototype and Back Again: Interactive Aspects of Cognitive Processing in Survey Interviews." Pp. 65-88 in *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, eds. Norbert Schwarz and Seymour Sudman. San Francisco: Jossey-Bass.

Schiffrin, Deborah. 1985. "Conversational Coherence: The Role of Well." *Language,* 61(3):640-667.

Wechsler, David. 1997. *Weschler Adult Intelligence Scale-Third Edition*. New York: The Psychological Corporation.

Yonker, James A., Robert M. Hauser, and Jeremy Freese. 2007. "The Dimensionality and Measurement of Cognitive Functioning at Age 65 in the Wisconsin Longitudinal Study." Center for Demography and Ecology Working Paper 2007-06, University of Wisconsin-Madison, Madison, WI. Unpublished manuscript.