Interviewers' willingness to spend time and effort on the survey: Workload and the contact process

Celine Wuyts and Geert Loosveldt

Centre for Sociological Research, KU Leuven

1. Introduction

The concept of "interviewer burden" suggests that if interviewers experience their tasks as more burdensome (e.g., because of a heavy workload, poor questionnaire design, long interviews, difficult to interview respondents) they adjust their behavior, taking shortcuts in order to reduce the expended time and effort on the survey ("interviewer satisficing"). Applied to face-to-face interviewers' activities during the contact procedure, this would imply that interviewers with a high burden may be less willing to make additional visits, may give up on difficult samples units prematurely, and may push for a quick decision instead of maintaining the interaction in a single or over multiple visits. Such interviewer satisficing behaviors in the contact procedure harm the effectiveness and efficiency of the fieldwork process, putting additional pressure on the quality-cost trade-off.

In this paper we focus on one aspect of interviewer burden: the workload within the survey, which varies over the fieldwork period. We assess the effect of this within-survey workload, on the outcomes of the contact process in the European Social Survey in Belgium.

Call history and assignment paradata allow the construction of more fine-grained measurements of interviewer workload over the fieldwork period than traditionally used. We will derive within-survey workload as a time-varying interviewer characteristic, measured as (1) the number of sample units for which interviewer activity has started and not yet ceased on any day *t*, and (2) the number of sample units for which interviewer activity is possible on any day *t*. Such operationalizations capture changes in the interviewers' workloads as the fieldwork progresses.

2. Data

We use data from Round 6 and 7 of the European Social Survey (ESS) in Belgium. The units in the Belgian gross samples, individual persons drawn from the National Register, were assigned to individual interviewers in several batches over the fieldwork period, and relatively large numbers of non-responding sample units were re-issued to other interviewers. Such a fieldwork strategy complicates the operationalization of within-survey workload over the fieldwork period, compared to a fieldwork strategy with all units assigned only once at the start of fieldwork.

In Round 6 (2012-2013), 1,869 interviews were administered by 155 interviewers from 3,267 sample units (response rate 58.7%), and in Round 7 (2014-2015), 1,769 interviews were administered by 150 interviewers from 3,267 sample units (response rate 57.0%). The same survey agency was contracted in both survey rounds and the two groups of interviewers partially overlap.

We construct several measurements for the main interviewer characteristic of interest, within-survey workload from the publicly available ESS call history data ("contact forms data") and additional paradata on the assignment of sample units to the interviewers. The following subsection offers a detailed description.

3. Operationalizing within-survey workload

Different ways of operationalizing workload carried by interviewers in the survey of interest itself (within-survey workload), can be considered. In previous studies, within-survey workload has been operationalized as the total number of interviews administered (e.g. Loosveldt and Beullens, 2013), or the total number of target addresses or respondents assigned to or worked on by each interviewer (e.g. Singer, Frankel and Glassman, 1983; Blom, 2012). Such aggregate measurements at the interviewer level are usually readily available or easy to derive but have their limitations. The interviewers' workload in a survey is not constant throughout the fieldwork period. Large variations can occur especially when cases are assigned in batches and/or re-assigned to different interviewers over the course of the fieldwork period. At any moment in time, we can expect interviewers to adjust their behavior in response to their actual workload, more so than to their past or anticipated workload. Operationalizing interviewer workload as a time-varying characteristic is therefore assumed to be more appropriate to explain interviewers' fieldwork effort and performance.

A useful alternative strategy would thus be to measure within-survey workload at different points in time over the fieldwork period. Such an operationalization captures variations in workload over time. The introduction of *time-varying* interviewer characteristics to explain changes in interviewer behavior and performance can be attributed to Olson and Peytchev (2007), who proposed using interview order to capture interviewers' increasing familiarity with the survey instrument over the fieldwork period. We analogously propose and assess two fine-grained, time-varying operationalizations of interviewer workload.

The measurement of within-survey workload over the fieldwork period hinges on the identification of the moments in time at which sample units enter and leave the interviewers' workloads. Two approaches appear reasonable. The first approach would consider a sample unit as being in an interviewer's workload as long as the interviewer is actively pursuing the unit in question. Each sample unit is assumed to enter the interviewer's workload on the date the interviewer first attempts to contact the sample unit and leaves the interviewer's workload on the date of the last attempt. Applying this first approach, within-survey workload at time *t* is measured as the number of 'active' sample units (sample units for which interviewer activity has started and not yet ceased) at that time. This approach requires detailed call history data, and a little more computation than the traditional aggregate measures.

The second approach would consider a sample unit as being in an interviewer's workload as long as the interviewer could *potentially* pursue the unit in question. Each sample unit is assumed to enter the interviewer's workload on the date the interviewer is assigned the sample unit and leaves the interviewer's workload when a final outcome code has been achieved or the sample unit is returned to the field office. Applying this second approach, within-survey workload at time *t* is measured as the number of 'potential' sample units (sample units for which interviewer activity is possible) at that time. This approach would be fairly straightforward if additional paradata on case assignments to interviewers and returns to the field office are available. This type of paradata is typically not available even if call history data is available, as is the case for the European Social Survey.

We therefore tested a procedure to derive a plausible approximation of interviewers' 'potential' case workloads. The key idea is to approximate the date of assignment and the date of return to the field office from observed interviewer activity in (implicit) sets of sample units. A stepwise procedure, progressively making more use of the available information, is used.

First, the unique combinations of sample units and interviewers are derived from the Belgian ESS call history data. They represent both the initial and re-issue assignments of sample units to interviewers for which the interviewer made at least one contact attempt (active assignments). Assignments of sample units which are never attempted (passive assignments) cannot be identified from the call history data. Table 1 summarizes the number of (active and passive) assignments and the number of sets of sample units derived at each of the following two steps.

Table 1. Identification of interviewers' sets of sample units

	ESS6	ESS7
Active assignments		
Number of (active) initial assignments (= number of sample units)	3,267	3,204
Number of (active) re-issue assignments	622	1,041
Total number of (active) assignments (Step 1)	3,889	4,245
Number of sets of (active) assignments based on overlapping interviewer activity (Step 2)	260	262
Number of sets of (active) assignments based on overlapping interviewer activity, geographical clusters, and allocation dates (Step 3)	583	749
Passive assignments		
Number of passive assignments (= number of sample units assigned to an interviewer without any activity)	160	321
Proportion of interviewers with passive assignments	12.26	25.83

In the second step, sample units for which the time span between the first and last contact attempt by the same interviewer overlaps are grouped in implicit assignment clusters using the R package `IRanges` (Lawrence et al., 2013). About 260 sets of sample units are identified in this way on the basis of overlap in interviewer activity for each survey round.

Each set of sample units is assumed to enter the interviewer's workload on the date the interviewer first attempt to contact a sample unit included in the set (approximation of set assignment date) and each set is assumed to leave the interviewer's workload on the date of the last attempt in the set (approximation of set finalization date). Respondents leave the interviewer's workload on the date of the completed interview. Sample units identified as deceased also leave the interviewer's workload.

Because by the Belgian ESS fieldwork strategy each sample unit is assigned to a single interviewer at any one point in time, we should not observe any sample units simultaneously allocated to multiple interviewers' workloads. We nonetheless do observe 5.1% and 13.2% of sample units simultaneously in multiple interviewers' workloads in Rounds 6 and 7, respectively. These high rates suggest that too many sample units may have been erroneously clumped together, inflating within-survey workload estimates.

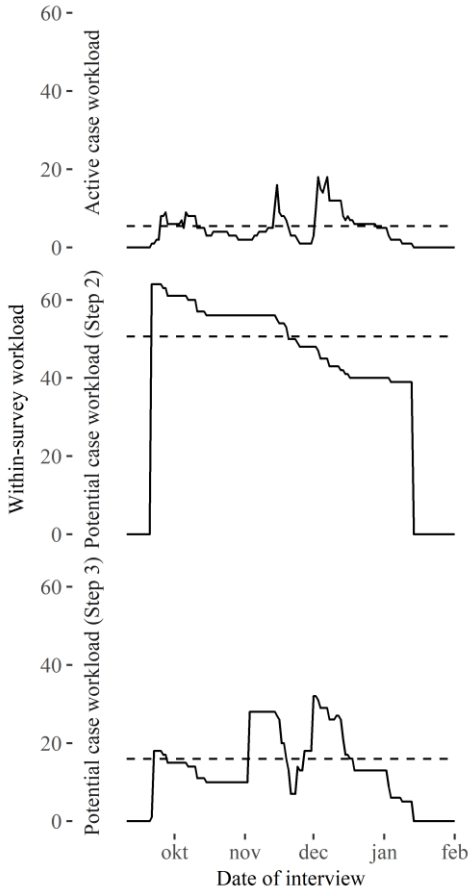
The third step aims to improve upon the grouping of sample units in sets and the determination of set assignment and finalization dates by taking into account the additional assignment data for ESS Belgium in Rounds 6 and 7. In each of the two survey rounds, about seven in ten sets of sample units previously identified on the basis of overlapping interviewer activity contain sample units with different assignment dates, and are therefore split up. After this operation, the number of identified sets of sample units more than doubles (Table 1).

Each set of sample units is assumed to enter the interviewer's workload on the date of assignment, unless the date of assignment is missing (N = 2 in Round 6, N = 6 in Round 7) or follows rather than precedes the first contact attempt (N = 83 in Round 6, N = 97 in Round 7), in which case the date the interviewer first attempt to contact a sample unit included in the set is taken. Because data on case returns to the field office are not available, each set is assumed to leave the interviewer's workload on the date of the last attempt in the set. As in the previous step, respondents leave the interviewer's workload on the date of the completed interview. Sample units identified as deceased also leave the interviewer's workload. We now observe only 0.1% and 2.1% of sample units in Rounds 6 and 7 simultaneously allocated to multiple interviewers' workloads.

Aggregating the number of sample units that have entered, and have not yet left, the workload of the interviewer at time *t* yields the interviewer's within-survey workload at that time *t*.

Figure 1 provides an illustration of how an interviewer's within-survey workload profile over the fieldwork period shifts as more of the available information is taken into account.

Figure 1. Illustration of an interviewers' time-varying within-survey workload



The interviewer presented was assigned a total of 64 sample units. The first panel tracks the number of sample units the interviewer started, and has not yet stopped, pursuing on each day over the fieldwork (active case workload). The second panel illustrates how grouping on the basis of overlapping interviewer activity alone (Step 2) may yield overestimated potential case workload estimates. In this case, all sample units of the interviewer are grouped together, erroneously suggesting that all were assigned at the start of fieldwork. The potential case workload estimates shown in the third panel, based on assignment sets and dates (Step 3) is much more realistic.

The time-varying measurements of within-survey workload can be used to model interviewer effort over the fieldwork period as illustrated in the following section.

4. Contact effort

The concept of interviewer burden suggests that heavier workloads would induce interviewers to expend less effort for each individual case in their workload. Making sufficient numbers of visits is one important component of the expected interviewer effort. We therefore test the hypothesis that when interviewers carry heavier workloads, their effort in terms of making visits is negatively affected.

We define **interviewer contact effort** on day ***t*** as the number of personal visits made by an interviewer *j* on day *t* of fieldwork, relative to the number of sample units in his (assigned case) workload on that day. Multilevel regression models with fieldwork days nested within interviewers are estimated using alternative operationalizations for within-survey workload at the level of fieldwork days and at the interviewer level to predict contact effort by interviewer *j* on day *t*.

In order to compare the predictive power of within-survey workload as a time-varying interviewer characteristic to within-survey workload as a fixed interviewer characteristic, we also construct the corresponding interviewer-level summary measures of the time-varying measurements of within-survey workload (Table 2). Taking the average of the time-varying measurements for each interviewer essentially corresponds to weighting (sets of) sample units by the amount of time they are in the interviewers' workloads. These new measures of interviewer workload at the interviewer level capture the active/potential case workload on a typical fieldwork day. The different operationalizations tested are presented in Table 2.

Table 2. Operationalizations of within-survey workload

	At the level of fieldwork days	**At the interviewer level**
Traditional operationalizations	-	Total active case workload = the total number of sample units worked on
New operationalizations	Active case workload on day *t* = the number of sample units for which interviewer activity has started and not yet ceased	Active case workload on a typical fieldwork day = the *average* number of sample units for which interviewer activity has started and not yet ceased
	Potential case workload on day *t* = the number of sample units for which interviewer activity is possible • derived from call history data (Step 2) • derived from assignment data (Step 3)	Potential case workload on a typical fieldwork day = the *average* number of sample units for which interviewer activity is possible

Table 3 shows the parameter estimates and model fit statistics for the four models with interviewer workload as a fixed interviewer characteristic and the three models with interviewer workload as a time-varying characteristic. The parameter estimate for within-survey workload is negative for all but one operationalization, indicating that interviewers carrying larger workloads tend to make fewer personal visits per case in their workload.

For both survey rounds, the models with the measures of *typical* potential case workload have a better fit than the model using the traditional *total* interviewer workload. Among the interviewer-level operationalizations, the potential case workload on a typical fieldwork day, as derived from the assignment data in Step 3, yields the best fit. The corresponding time-varying operationalization, the daily potential case workload, as derived from the assignment data in Step 3, yields an even better fit in both survey rounds. The potential case workload, as derived from the call history data in Step 2, has lower predictive power than the potential case workload as derived from the assignment data in Step 3, both when measured for a typical fieldwork day at the interviewer level and when measured on the level of fieldwork days.

Interviewers' daily active case workload yields a very good fit in both rounds but a parameter estimate in the opposite direction from what would be expected. The observation that larger numbers of active cases are *positively* associated with the number of personal visits that are made is probably an artefact of this particular operationalization, i.e. the measurement itself depends heavily on the

distribution of visits made by the interviewers. Note that this problem appears mitigated for the corresponding interviewer-level *typical* active case workload, but this operationalization does not yield a good model fit.

Table 3. Parameter estimates for alternative operationalizations for within-survey workload (in tens of cases) to predict interviewers' **daily contact effort**, and model fit statistics relative to null model

	ESS6		ESS7	
Within-survey workload operationalization	**Est.**	**ΔAIC**	**Est.**	**ΔAIC**
Fixed interviewer characteristic				
Total active case workload	-0.0113 *	-2.50	-0.0126 **	-7.11
Active case workload on a typical fieldwork day	-0.0538	0.29	-0.0945 *	-3.68
Potential case workload (derived from call history data, Step 2) on a typical fieldwork day	-0.0409 ***	-14.91	-0.0438 ***	-34.63
Potential case workload (derived from assignment data, Step 3) on a typical fieldwork day	-0.1557 ***	-46.15	-0.1571 ***	-71.46
Time-varying interviewer characteristic				
Active case workload on day *t* [a]	0.1890 ***	-247.29	0.1135 ***	-94.85
Potential case workload (derived from call history data, Step 2) on day *t*	-0.0225 ***	-11.21	-0.0391 ***	-62.60
Potential case workload (derived from assignment data, Step 3) on day *t*	-0.0778 ***	-90.16	-0.1036 ***	-226.69
N	2722		2951	

† p < .10; * $p < .05$; ** $p < .01$; *** $p < .001$

Because potential case workload, especially when derived from the assignment data in Step 3, appears to be the most adequate operationalization of within-survey workload, we will focus our attention on this operationalization, relative to its aggregate counterpart at the interviewer level and the traditional operationalization of interviewer workload as total active case workload in the following sections.

5. Interviewer contact and cooperation rates

The obvious consequence of reduced effort in contacting and recruiting target respondents when interviewers carry heavier workloads, as described in the previous section, is that the outcomes of the fieldwork process will be negatively affected by heavy workloads.

The effect of workload on interviewers' fieldwork performance is assessed by estimating regression models using alternative operationalizations for within-survey workload to predict interviewer contact and cooperation rates at three levels of aggregation (fieldwork days, fieldwork weeks, and the overall fieldwork period).

Contact and (conditional) cooperation rates for each interviewer were derived from the ESS call history data. The **interviewer contact rate** (on day *t*, in week *s*, over the entire fieldwork) is defined as the relative number of sample units visited by interviewer *j* (on day *t*, in week *s*, over the entire fieldwork) with which contact with the target respondent was successfully made. The **interviewer cooperation rate** (on day *t*, in week *s*, over the entire fieldwork) is defined as the relative number of target respondents contacted by interviewer *j* (on day *t*, in week *s*, over the entire fieldwork) from which an interview was successfully completed.

Parameter estimates and model fit statistics for interviewers' daily, weekly and overall contact rates are presented in Table 4, Table 5 and Table 6, respectively. At least when statistically significant, the parameter estimates are negative as expected.

We observe that in either survey round and at either level of aggregation, the traditional operationalization at the interviewer level of total active case workload has no predictive power with regard to interviewers' contact rates.

The results for the two potential case workload operationalizations are somewhat different for the two survey rounds. At the level of fieldwork days, the potential case workload operationalizations, both at the interviewer level and at the level of fieldwork days yield some improvement in model fit relative to the null model in Round 6, but have no predictive power in Round 7. Moving up to the level of fieldwork weeks, the expected improvements in model fit from a *typical* potential case workload relative to the *total* case workload, and from the daily potential case workload relative to typical potential case workload become more pronounced in both survey rounds. At the interviewer level, typical potential case workload remains negatively associated with contact rates in both rounds, but in Round 7 the estimate is not statistically significant at the 5% and the predictive power is low.

Table 4. Parameter estimates for alternative operationalizations for within-survey workload (in tens of cases) to predict interviewers' **daily contact rate**, and model fit statistics relative to null model

	ESS6		ESS7	
Within-survey workload operationalization	**Est.**	**ΔAIC**	**Est.**	**ΔAIC**
Fixed interviewer characteristic				
Total active case workload	0.0058	1.59	-0.0015	1.95
Potential case workload (derived from assignment data, Step 3) on a typical fieldwork day	-0.1187 **	-6.40	-0.0547	-0.58
Time-varying interviewer characteristic				
Potential case workload (derived from assignment data, Step 3) on day *t*	-0.0400 **	-6.60	-0.0003	2.00
N	2722		2951	

† p < .10; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 5. Parameter estimates for alternative operationalizations for within-survey workload (in tens of cases) to predict interviewers' **weekly contact rate**, and model fit statistics relative to null model

	ESS6		ESS7	
Within-survey workload operationalization	**Est.**	**ΔAIC**	**Est.**	**ΔAIC**
Fixed interviewer characteristic				
Total active case workload	0.0018	1.96	-0.0033	1.78
Potential case workload (derived from assignment data, Step 3) on a typical fieldwork day	-0.1388 **	-8.29	-0.0670 †	-1.77
Time-varying interviewer characteristic				
Potential case workload (derived from assignment data, Step 3) on a typical fieldwork day in week *s*	-0.1244 ***	-48.46	-0.0563 ***	-12.96
N	1136		1265	

† p < .10; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 6. Parameter estimates for alternative operationalizations for within-survey workload (in tens of cases) to predict interviewers' **overall contact rate**, and model fit statistics relative to null model

	ESS6		ESS7	
Within-survey workload operationalization	**Est.**	**ΔAIC**	**Est.**	**ΔAIC**
Fixed interviewer characteristic				
Total active case workload	-0.0012	1.97	0.0009	1.98
Potential case workload (derived from assignment data, Step 3) on a typical fieldwork day	-0.1146 ***	-13.78	-0.0522 †	-1.60
N	155		151	

† p < .10; * $p < .05$; ** $p < .01$; *** $p < .001$

Parameter estimates and model fit statistics for interviewers' daily, weekly and overall cooperation rates are presented in Table 7, Table 8 and Table 9, respectively. As for interviewers' contact rates, the interviewer workload parameter estimates are negative as expected, at least when statistically significant. The results suggest that when interviewers carry larger workloads they tend to be less successful in convincing contacted target respondents to participate in the survey.

Also as for interviewers' contact rates, in either survey round and at either level of aggregation, the traditional operationalization of total active case workload has no predictive power with regard to interviewers' cooperation rates. Improvements in model fit from using *typical* potential case workload, relative to the *total* case workload, and from the time-varying potential case workload relative to typical potential case workload are observed for both survey rounds and each level of aggregation.

Table 7. Parameter estimates for alternative operationalizations for within-survey workload (in tens of cases) to predict interviewers' **daily cooperation rate**, and model fit statistics relative to null model

	ESS6		**ESS7**	
Within-survey workload operationalization	**Est.**	**ΔAIC**	**Est.**	**ΔAIC**
Fixed interviewer characteristic				
Total active case workload	0.0099	0.18	-0.0003	2.00
Potential case workload (derived from assignment data, Step 3) on a typical fieldwork day	-0.0859 *	-3.97	-0.0572 *	-1.99
Time-varying interviewer characteristic				
Potential case workload (derived from assignment data, Step 3) on day *t*	-0.0809 ***	-21.80	-0.0433 **	-7.88
N	2177		2220	

† p < .10; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 8. Parameter estimates for alternative operationalizations for within-survey workload (in tens of cases) to predict interviewers' **weekly cooperation rate**, and model fit statistics relative to null model

	ESS6		**ESS7**	
Within-survey workload operationalization	**Est.**	**ΔAIC**	**Est.**	**ΔAIC**
Fixed interviewer characteristic				
Total active case workload	0.0104	0.27	-0.0019	1.91
Potential case workload (derived from assignment data, Step 3) on a typical fieldwork day	-0.1108 **	-6.50	-0.0580 †	-1.58
Time-varying interviewer characteristic				
Potential case workload (derived from assignment data, Step 3) on a typical fieldwork day in week *s*	-0.1319 ***	-39.75	-0.0734 ***	-15.93
N	1036		1098	

† p < .10; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 9. Parameter estimates for alternative operationalizations for within-survey workload (in tens of cases) to predict interviewers' **overall cooperation rate**, and model fit statistics relative to null model

	ESS6		**ESS7**	
Within-survey workload operationalization	**Est.**	**ΔAIC**	**Est.**	**ΔAIC**
Fixed interviewer characteristic				
Total active case workload	0.0036	1.88	0.0007	1.99
Potential case workload (derived from assignment data, Step 3) on a typical fieldwork day	-0.1859 ***	-18.10	-0.0988 **	-4.88
N	155		151	

† p < .10; * $p < .05$; ** $p < .01$; *** $p < .001$

Conclusion

The optimal performance of the contact and recruitment task do appear to be at risk from heavy interviewer workloads inducing interviewers to expend less time and effort for individual units. The initial evidence suggests that when interviewers carry heavier workloads they tend to make fewer visits to each case in their workload, resulting in lower contact rates achieved when outcomes are aggregated at least at the level of fieldwork weeks, as well as lower cooperation rates.

Interviewer characteristics are usually assumed fixed over the fieldwork period. In this paper we show that since interviewer workload can vary strongly over the fieldwork period, a more fine-grained operationalization can be useful to explain differences in interviewers' fieldwork performance and the underlying mechanisms. Among the operationalizations at the interviewer level, *typical* workload, as derived from the detailed, daily workload measurement, consistently outperforms the traditional operationalization of *total* case workload. The detailed, time-varying workload operationalization itself in turn fairly consistently outperforms the interviewer-level *typical* workload that is derived from it. The results are especially pronounced for interviewers' contact effort, more so than for interviewers' contact and cooperation rates. That the link between workload and contact effort is stronger than the link between workload and fieldwork outcomes, even at low levels of aggregation, may be explained by effort being completely within the control of the interviewers, and thus more directly affected by interviewer burden.

References

Blom, A. G. (2012), "Explaining cross-country differences in survey contact rates: Application of decomposition methods: cross-country differences in survey contact rates," Journal of the Royal Statistical Society, *Statistics in Society*, 175, 217–242. https://doi.org/10.1111/j.1467-985X.2011.01006.x.

Botman, S. L., and Thornberry, O. T. (1992), "Survey design features correlates of nonresponse," in *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 309–314.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., and Carey, V. J. (2013), "Software for Computing and Annotating Genomic Ranges," *PLoS Computational Biology*, 9, e1003118. DOI: 10.1371/journal.pcbi.1003118.

Loosveldt, G., and Beullens, K. (2013), "'How long will it take?' An analysis of interview length in the fifth round of the European Social Survey," *Survey Research Methods*, 7, 69–78.

Olson, K., and Peytchev, A. (2007), "Effect of interviewer experience on interview pace and interviewer attitudes," *Public Opinion Quarterly*, 71, 273–286. DOI: 10.1093/poq/nfm007.

Singer, E., Frankel, M. R., and Glassman, M. B. (1983), "The effect of interviewer characteristics and expectations on response," *Public Opinion Quarterly*, 47, 68. https://doi.org/10.1086/268767.