

University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Sociology Department, Faculty Publications

Sociology, Department of

2017

Examining Changes of Interview Length Over the Course of the Field Period


Antje Kirchner

University of Nebraska-Lincoln, akirchner@rti.org

Kristen Olson

University of Nebraska-Lincoln, kolson5@unl.edu

Follow this and additional works at: <http://digitalcommons.unl.edu/sociologyfacpub>

 Part of the [Family, Life Course, and Society Commons](#), and the [Social Psychology and Interaction Commons](#)

Kirchner, Antje and Olson, Kristen, "Examining Changes of Interview Length Over the Course of the Field Period" (2017). *Sociology Department, Faculty Publications*. 511.

<http://digitalcommons.unl.edu/sociologyfacpub/511>

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Sociology Department, Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

KIRCHNER & OLSON IN *J. OF SURVEY STATISTICS AND METHODOLOGY*
CHANGES OF INTERVIEW LENGTH OVER THE FIELD PERIOD

Published in *Journal of Survey Statistics and Methodology* 5 (2017), pp. 84–108.
doi: 10.1093/jssam/smw031

Copyright © 2016 Antje Kirchner and Kristen Olson. Published by Oxford University Press on behalf of the American Association for Public Opinion Research. Used by permission.

Published online 24 December 2016.

This material is based upon work supported by the National Science Foundation under Grant No. SES-1132015. Opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Earlier versions of this paper were presented at the American Association for Public Opinion Research annual meeting, May 2014, Anaheim, CA, and the Joint Statistical Meetings, August 6, 2014, Boston, MA, USA.

Examining Changes of Interview Length Over the Course of the Field Period

Antje Kirchner and Kristen Olson

Antje Kirchner, Research Survey Methodologist and Adjunct Research Assistant Professor, is with the Survey Research Division, RTI International, Research Triangle Park and the Department of Sociology, University of Nebraska–Lincoln.

Kristen Olson, Associate Professor, is with the Department of Sociology, University of Nebraska–Lincoln.

Corresponding author — Antje Kirchner, Survey Research Division, RTI International, Research Triangle Park, NC; E-mail: akirchner@rti.org.

Abstract

It is well established that interviewers learn behaviors both during training and on the job. How this learning occurs has received surprisingly little empirical attention: Is it driven by the interviewer herself or by the respondents she interviews? There are two competing hypotheses about what happens during field data collection: (1) interviewers learn behaviors from their previous interviews, and thus change their behavior in reaction to the behaviors previously encountered; and (2) interviewers encounter different types of and, especially, less cooperative respondents (i.e., nonresponse propensity affecting the measurement error situation), leading to changes in interview behaviors over the course of the field period. We refer to these hypotheses as the experience and response propensity hypotheses, respectively. This paper examines the relationship between proxy indicators for the experience and response propensity hypotheses on interview length using data and paradata from two telephone surveys.

Our results indicate that both interviewer-driven experience and respondent-driven response propensity are associated with the length of interview. While general interviewing experience is nonsignificant, within-study experience decreases interview length significantly, even when accounting for changes in sample composition. Interviewers with higher cooperation rates have significantly shorter interviews in study one; however, this effect is mediated by the number of words spoken by the interviewer. We find that older respondents and male respondents have longer interviews

despite controlling for the number of words spoken, as do respondents who complete the survey at first contact. Not surprisingly, interviews are significantly longer the more words interviewers and respondents speak.

Keywords: Interviewer experience, Interviewer learning, Response propensity, Paradata

1. Introduction

Experienced and inexperienced interviewers differ in how they implement survey protocols and interact with respondents (Cleary, Mechanic, and Weiss 1981; Fowler and Mangione 1990). Four broad types of experience are typically distinguished (Schaeffer, Dykema, and Maynard 2010): life-time experience as an interviewer, experience or tenure within a survey company, study-specific experience, and the number of different surveys worked (Groves and Couper 1998; Olson and Peytchev 2007; Durrant, Groves, Staetsky, and Steele 2010; Olson and Bilgen 2011; Loosveldt and Beullens 2013a, 2013b; Böhme and Stöhr 2014). Interviewers are hypothesized to change some of the behaviors they had been taught during training as they conduct repeated interviews.

In particular, in face-to-face (F2F) surveys, interview length decreases both over the course of a study and with more overall experience. Olson and Peytchev (2007) find that inexperienced interviewers initially take significantly longer than experienced interviewers, but interview length decreases much more quickly for these interviewers over the field period. Recent studies conducted by Loosveldt and Beullens (2013a, 2013b) support the general finding of a decrease in F2F interview length with increasing survey-specific experience, although the authors do not account for general interviewing experience. Böhme and Stöhr (2014) do not find this effect for survey-specific experience, perhaps because they account for more interviewer demographics. All of these studies focus on face-to-face surveys; we do not know whether this same effect holds in centralized telephone surveys.

Previous studies attribute the decline in interview length to interviewers gaining experience and practicing interviewing behaviors. That is, the *experience hypothesis* posits that interviewers learn from prior experiences and thus alter the measurement situation, either overall or within a particular study (Olson and Peytchev 2007). Interviewers with more overall experience usually have established routines in administering a survey. Experienced interviewers may also lose their standardized behaviors from training over time and may be less likely to be “retrained” on generalized interviewing techniques than their less experienced counterparts. Thus, we expect that experienced interviewers will have shorter interview lengths than inexperienced interviewers.

Study-specific experience also operates through similar mechanisms. First, as with overall experience, interviewers may read questions more fluently and smoothly and anticipate potential problems before they occur as they become more practiced during the survey itself, resulting in shorter interviews over the field period. Second, interviewers might become less attentive over time, potentially altering or skipping question wording. These changes are likely to occur during the beginning of the field period and lessen as

data collection progresses (e.g., Olson and Peytchev 2007). We also expect that inexperienced interviewers will have the largest learning effects over the course of the field period.

A previously unexplored type of experience results from interactions during the recruitment stage rather than the measurement stage. Interviewers who are more skilled with “tailoring” (nonstandardized recruitment techniques) might revert to those more often during the measurement situation (Brunton-Smith, Sturgis, and Williams 2012; Snijkers, Hox, and de Leeuw 1999). There is initial evidence that interviewer cooperation rates are also associated with different types of interviewer behaviors during the interview, especially related to display of confidence such as fewer disfluent paralinguistic cues and less stuttering (Olson, Kirchner, and Smyth 2016). Thus, we expect that interviewers with higher cooperation rates, reflecting more successful experiences at recruitment, will have shorter interviews.

An alternative hypothesis, the *response propensity hypothesis*, states that interviewers encounter different types of respondents as fieldwork progresses. This change could manifest either as a change in sample composition, contactability, or cooperation, the respondent-interviewer interaction, or all three.

The *compositional aspect* of the response propensity hypothesis argues that respondents with different characteristics are recruited over the course of the field period. Older and less educated respondents will have longer interviews due to less working memory capacity (Yan and Tourangeau 2008; Couper and Kreuter 2013; Olson and Smyth 2015). Competing time demands or distractions that result from family or work circumstances could also shorten interview length (Couper and Kreuter 2013). Prior studies on interview length have accounted for age and education (Olson and Peytchev 2007; Loosveldt and Bullens 2013a, 2013b) and one study (on acquiescence) has accounted for sex and race of respondent (Olson and Bilgen 2011), but none have accounted for proxy measures of time demands. To the extent that faster respondents are recruited later in the interviewer’s pool, then sample composition rather than learning may be the reason for decreases in interview length.

The second part of the response propensity hypothesis focuses on a change in the combination of contactable versus away-from-home and cooperative versus reluctant cases throughout the field period. Respondents with a lower propensity to respond may be less engaged in forming a response or more generally provide lower quality data by providing more “don’t know” responses or acquiescent responses (Olson 2013). Finally, interviewers might change their behavior as they encounter more reluctant respondents or engage in probing less often in order to keep reluctant respondents from breaking off the survey interview.

A third component of the response propensity hypothesis has to do with the complexity of the interviewer and respondent *interaction* itself. The length of interview is driven by respondent interactions with the interviewer. If the components of this interaction change over the course of the field period, so will the interview length. For example, later respondents may speed through the interview or engage less with the interviewer. Interviewers might adjust their behavior accordingly, for example, by increasing their interviewing pace and reducing the

amount of interaction. This should lead to a decrease in interview length over the course of the field period to the extent that less interactive respondents are recruited later in the field period.

Prior studies attribute the decrease in interview length over the course of the data collection period primarily to growing survey-specific interviewer experience. The main contribution of this paper is to empirically investigate this hypothesis in a different mode (telephone) while examining alternative explanations that could result in the same observed outcomes. More specifically, we disentangle two competing hypotheses: *experience* (driven by interviewer influences) and *response propensity* (driven by respondent influences). We expand on previous work by examining survey data and paradata from two independent telephone surveys (rather than face-to-face surveys), different aspects of interviewer experience, including successful experience in recruiting respondents, and a variety of respondent measures of response propensity and sample composition. We use multilevel models to account for clustering of respondents due to interviewers.

2. Data and Methods

2.1 *Work and Leisure Today 1*

We use data and paradata from the Work and Leisure Today (WLT1) survey, a computer-assisted telephone survey collected by AbtSRBI in August 2013. AbtSRBI introduced software (VOXCO) prior to data collection that was new to all interviewers. Overall, 450 interviews were completed by a total of 22 interviewers (AAPOR RR3=6.3%) (see Olson and Smyth 2015 for details). Three interviewers with fewer than 10 interviews were removed from the analyses due to lack of learning opportunities (Olson and Peytchev 2007). This results in an analytic sample of 435 interviews conducted by 19 interviewers.

The WLT1 uses a random digit dialing (RDD) sample of US landlines only. At the time of this study, 39.1% of adults did not have landline telephone service (Blumberg and Luke 2014). This raises coverage issues: compared to the 2013 ACS three-year estimates at the time of the study (estimates derived for adults [Census Bureau 2016a]), the overall WLT1 sample contains more females (64.4% versus the national estimate of 51.4%), is older (79.4% aged 50+ versus the national estimate of 43.4%), and is more likely to be white (87.3% white versus the national estimate of 75.7%).

2.2 *Work and Leisure Today 2*

The second study is the Work and Leisure Today 2 (WLT2) survey conducted in September 2015. Data were collected by AbtSRBI with identical data collection protocols and a similar questionnaire to the WLT1, but with an independent sample of phone numbers (that is, WLT2 was not a follow-up wave to WLT1). A total of 902 interviews were completed in WLT2. Removing interviewers with fewer than 10 interviews results in an analytic sample of 899 interviews conducted by 26 interviewers.

The WLT2 remedies the two major limitations of the original WLT study, that is, representativeness and interviewer familiarity with the interviewing software. The WLT2 is a dual-frame survey with both landline and cell phone numbers (landline=451, AAPOR RR3=9.4%;

cell phone=451, AAPOR RR3=7.1%). The WLT2 sample is more representative of the adult population than WLT1. While the overall sample is still older compared to 2014 ACS five-year estimates (64.5% aged 50+ versus the national estimate of 43.4%), the WLT2 is more representative regarding gender (51.8% versus the national estimate of 51.4%) and race (79.6% versus the national estimate of 75.6%) (estimates derived for adults [Census Bureau 2016b]).

2.3 Modeling Interview Length

Interview length, the dependent variable, is derived from paradata and measured in minutes from the administration of the first survey question to the last answer. Outlier observations were trimmed and replaced with values at the 1st and the 99th percentile (WLT1 mean: 12.7minutes; WLT2 mean: 13.4minutes) (Ratcliff 1993; Yan and Tourangeau 2008).

2.4 Interviewer Experience (H1)

We operationalize within-survey experience with an ordinal counter, ranging from 1 for the first interview completed by each interviewer to maximum of 27 in the WLT1 and 79 in the WLT2 for the last interview. We use the natural logarithm of interview order to account for a nonlinear learning curve (Olson and Peytchev 2007). We include an indicator of overall interviewer experience, operationalized using the time employed at AbtSRBI, where 1=1+years of experience and 0=no years of experience (Olson and Bilgen 2011). We also expect that the learning curve is steeper for more inexperienced interviewers. Thus, an interaction term of general interviewing experience and study-specific experience is included in the models.

As learning might occur during survey recruitment, we include a measure of experience in obtaining respondent participation. Interviewer-level cooperation rate (conditional on contact) has an average of 6.8% in the WLT1 and 11.8% in the WLT2. We expect that interviewers who are more successful in recruiting respondents will have a shorter interview length.

2.5 Modeling Response Propensity–Composition Hypothesis (H2a)

The sociodemographic composition of the sample may change systematically over the course of the field period. Thus, we include the respondent demographics of gender, age, education, and race. We include measures of time demands and potential distractions including employment, household income, household size, being a parent, and volunteering. Sample composition differs for early and late respondents despite the generally low response rates (see appendix A.1).

2.6 Modeling Response Propensity–Contactability and Cooperation Hypothesis (H2b)

The second aspect of the response propensity hypothesis examines respondent *contactability and cooperation*. Previous studies associate higher recruitment effort with a higher risk of poor data quality, with the item nonresponse rate being the most sensitive data quality

indicator for this association (Olson 2013). We include the item nonresponse rate across the entire survey, whether the respondent ever refused (as a soft refusal) to participate in the survey, and whether the interview was completed at first contact as reluctance measures. We measure contactability using the number of call attempts. The call window when the interview was successfully completed¹ captures respondent-specific characteristics that cannot be accounted for with our data but that are related to both contactability and reluctance. We expect that interviews completed on a weekday evening (5:00 P.M. to 11:00 P.M.) will be shorter than those completed on the weekend because of increased time constraints on weekdays. We anticipate that respondents who are harder to contact and who are more reluctant will have shorter interviews.

2.7 Modeling Response Propensity–Interaction (H2c)

We include the number of words spoken (separately) by the respondent and the interviewer as indirect measures of the interaction (divided by 100; censored at the 1st and 99th percentile). The word count is obtained from interview transcripts for WLT1 (unavailable for WLT2). Naturally, the number of words spoken during the interview will be related to interview length. However, the mechanism affecting interview length differs by actor. Interviewers are trained to follow interviewing protocols, with deviations from interviewing scripts largely in reaction to the respondent. Respondents have more flexibility in answering questions and engaging with the interviewer. If respondents who are more engaged talk more and are recruited early in the field period, we would expect to see longer interviews during the earlier interviews. To the extent that interviewers adjust their own behaviors to less talkative respondents (e.g., tailoring and maintaining interaction), this would lead to a decrease in the number of words spoken by the interviewer over the field period and a decrease in interview length.

2.8 Control Variables

All models control for the number of questions asked, the number of answer changes,² computer usage (controlling for skip patterns), and interviewer race and gender. No other interviewer characteristics were available. Following Stokes and Yeh (1988), we control for the interviewer’s typical shift, operationalized as being more likely to work on weekday evenings versus weekends.³ The WLT2 controls for “version,” as respondents were randomly allocated to one of two questionnaire versions.

All continuous independent variables are grand mean centered (Raudenbush and Byrk 2002). Table 1 provides an overview of all interviewer and respondent characteristics.

3. Methods

We use two-level hierarchical linear models with random intercepts to model interview length accounting for the clustering of respondents (level 1) within interviewers (level 2) (e.g., Hox 1994, 2010; Snijders and Bosker 2012; Raudenbush and Byrk 2002), estimated using Stata 14.1.

We examine each of the hypotheses separately in individual models for both studies and then evaluate their joint effects. For reasons of parsimony, we only report the full models (see appendix A.2 for all models). Given the comparatively large ICC that reduces the effective sample size and the small number of respondents and interviewers, we report all significant effects using an alpha level of 0.10.

4. Findings

Table 2 provides the variance components and the intraclass correlation coefficients (ICC) for the null models. Interviewers account for 21% of the total variance in interview length in WLT1 and 25% of the variance in length in WLT2.

Table 3 displays the regression coefficients and standard errors for the full models predicting interview length for WLT1 and WLT2. In order to compare the results between WLT1 and WLT2, we present the results for WLT1 with and without the word count variables.

5. Interviewer Experience

As expected, we observe a significant decrease in interview length in both studies over time. However, this effect is the same for both experienced and inexperienced interviewers.⁴ Figure 1 displays the effect of study-specific experience from the full model (WLT1 and WLT2). On average, interview length decreases by around two and a half minutes between the first and the 20th interview in WLT1 and by around two minutes in WLT2. Thus, the effects of within-study experience in WLT1 are not confounded by gaining practice with the new software but instead are due to study-specific learning.

We would expect to see a *weakening* in the coefficient for within-survey experience if our measures of response propensity explained this learning effect. The exact opposite is the case. As shown in Appendix A.2, the coefficient for interview order *increases* after including the sample composition and response propensity variables in the model. Including word counts, however, reduces this effect to its original magnitude (WLT1 incl. H2c).

Additionally, as the measure of experience at recruitment, WLT1 interviewers with higher cooperation rates have shorter interview lengths than interviewers with lower cooperation rates. This effect is fully absorbed by the number of words spoken by the interviewer, providing indirect support for interviewer tailoring (analyses available upon request). We do not see the cooperation rate effect in WLT2.

We now examine our second hypothesis, that is, response propensity that manifests in differential *composition*. We see that males, older respondents, racial minorities, and respondents living in larger households have longer interviews than females, younger respondents, white respondents, and respondents living in smaller households. Many of these effects hold even when including the number of words spoken by the respondent. This suggests the difference in interview length for females, younger respondents, and those living in smaller households cannot be simply attributed to saying more or fewer words. Measures of socioeconomic status, including education, employment, and income have inconsistent effects across the two studies (see table 3).

The *contactability and cooperation* component of the response propensity hypothesis posited that reluctant and hard-to-contact respondents will have shorter interviews. Contrary to our expectations, interviews with a higher item nonresponse rate take significantly *longer* to complete than those with a lower item nonresponse rate, potentially because “don’t know” or “refusal” responses are triggering interviewer probing behaviors. We see that this effect is fully absorbed when including word counts, strengthening this argument. Interviews completed at first contact take longer in WLT2. Calls that were completed during a weekend day in WLT1 were significantly shorter than a weekday evening by a minute in length. This effect is nonsignificant in WLT2; however, when the WLT2 analyses are restricted to the landline frame only, interviews completed on a weekend day are significantly shorter than those on a weekday evening in WLT2 (results not shown). The two most commonly used indicators of response propensity—the number of call attempts and an indicator for refusal conversion—were not associated with interview length in either survey.

The results for the *interaction* hypothesis in WLT1 are not surprising. More words spoken by the respondent or the interviewer are associated with significantly longer interviews. These indicators mediate the relationship of interview length and some variables (e.g., cooperation rate and item nonresponse rate) whereas other variables remain unaffected (e.g., gender and age). Including these indicators significantly decreases respondent-level variance.

6. Discussion

The goal of this paper was to investigate changes in interview length by examining two major hypotheses—experience and response propensity—using two independent studies. Replicating prior research, our results suggest that within-survey experience affects interview length—later interviews were about two minutes shorter than earlier interviews. Contrary to other studies, this learning effect did not differ for more or less experienced interviewers. It is possible that the close monitoring of telephone interviewers negates some of the overall experience effects observed in face-to-face surveys. Accounting for sample composition, contactability and cooperation, and one measure of interaction did not alter any of those results. Additionally, in WLT1, interviewers with higher cooperation rates have shorter interviews, although this effect did not replicate in WLT2 and is fully absorbed when including indicators of the respondent-interviewer interaction. More successful interviewers may also be more fluent in their delivery, shortening the length of the interview (Olson, Kirchner, and Smyth 2016).

Respondent-level response propensity measures provide mixed support for the response propensity hypothesis for interview length. In both studies, respondents with higher item nonresponse rates had longer interviews, not shorter interviews, suggesting that either interviewers or respondents (or both) act differently when a “don’t know” or “refusal” response is provided. This suggestion is supported by the fact that item nonresponse rates are no longer significant when we control for interactional indicators. Although there was no clear association between the number of call attempts or refusal conversion

and interview length, respondents who completed the interview at first contact had longer interviews in WLT2. Additionally, respondents interviewed on weekend days complete the interview more quickly in WLT1 (and landline respondents in WLT2). Interviewer training could encourage interviewers to motivate and engage respondents, especially during certain interviewer shifts or at later call attempts.

As with previous research in face-to-face surveys, we found that interviews get shorter over the field period in two telephone surveys, and this held even after accounting for a much broader array of potential contributors than any prior study. However, without information on what is actually happening in the interview—that is, more information on the interaction between the interviewer and respondent—the implications of these results for data quality are less clear. Do interviewers simply get better or sloppier? Do certain good or bad behaviors go away? Future research should address these issues. If interviewers actually “learn” in a positive way, more training interviews could be conducted before the actual fieldwork starts. On the other hand, if interviewers are losing desired standardized behaviors, a stronger focus on retraining and supervision during the field period may be needed.

Finally, interview length is only an indirect indicator of data quality (e.g., Olson and Parkhurst 2013; Yan and Olson 2013). Future research should also investigate the effects of variables capturing learning and response propensity on more direct indicators of data quality, such as rounding or response styles.

Overall, our results suggest that telephone interviewers do change some behaviors over time. The composition, contactability and cooperation, and interaction mechanisms do not override the changes due to experience, but instead contribute to them independently. Future research will examine more interviewer and respondent behaviors to further understand drivers of interview length.

Notes

1. Respondents who partially completed the interview and were reattempted were assigned the date and time of the first part of the interview. No interviews were completed during a weekday.
2. This indicator captures the number of recorded answers that were changed at any point during the interview.
3. As West and Olson (2010) found, interviewers crossed shifts. We identified interviewers based on working at least 90.3% of calls on weekday evenings (the 75th percentile) as typically working the weekday evening shift.
4. Appendix A.3 excludes the interaction effect of interview order and interviewer experience.

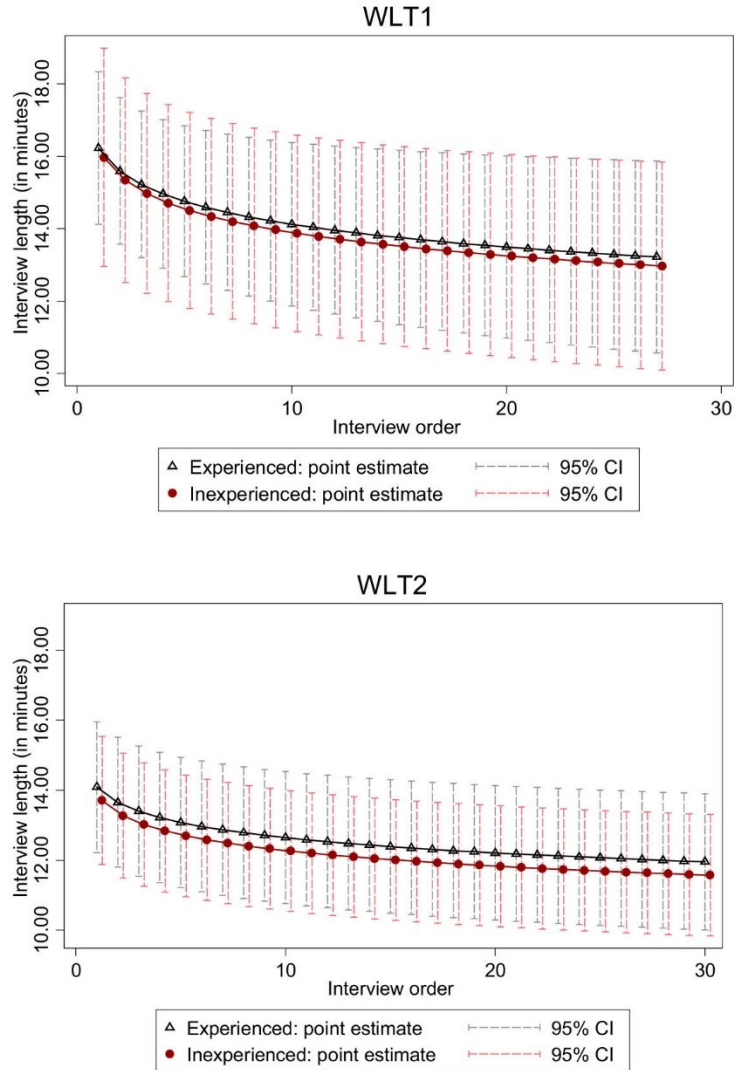


Figure 1. Expected interview length by overall and within-study experience

References

- The American Association for Public Opinion Research (AAPOR) (2015), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (8th ed.), AAPOR. [http://www.aapor.org/Standards-Ethics/Standard-Definitions-\(1\).aspx](http://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx)
- Blumberg, S. J., and J. V. Luke (2014), "Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, July–December 2013," U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, available at <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201407.pdf>
- Böhme, M., and T. Stöhr (2014), "Household Interview Duration Analysis in CAPI Survey Management," *Field Methods*, 26(4), 390–405. DOI: 10.1177/1525822X14528450
- Brunton-Smith, I., P. Sturgis, and J. Williams (2012), "Is Success on the Doorstep Correlated with the Magnitude of the Interviewer Design Effect?" *Public Opinion Quarterly*, 76, 265–286.

- Census Bureau (2016a), "2011-2013 American Community Survey 3-Year Estimates," Table ACS_13_3YR_B01001 and ACS_13_3YR_B01001A; http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_13_3YR_B01001&prodType=table and http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_13_3YR_B01001A&prodType=table
- Census Bureau (2016b), "2010-2014 American Community Survey 5-Year Estimates," Table ACS_14_5YR_B01001 and ACS_14_5YR_B01001A; http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_14_5YR_B01001&prodType=table and http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_14_5YR_B01001A&prodType=table
- Cleary, P. D., D. Mechanic, and N. Weiss (1981), "The Effect of Interviewer Characteristics on Responses to a Mental Health Interview," *Journal of Health and Social Behavior*, 22, 183–193.
- Couper, M. P., and F. Kreuter (2013), "Using Paradata to Explore Item Level Response Times in Surveys," *Journal of the Royal Statistical Society A*, 178, 271–286.
- Durrant, G. B., R. M. Groves, L. Staetsky, and F. Steele (2010), "Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys," *Public Opinion Quarterly*, 74, 1–36.
- Fowler, F. J. Jr., and T. W. Mangione (1990), *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*, Newbury Park: Sage.
- Groves, R. M., and M. P. Couper (1998), *Nonresponse in Household Interview Surveys*, New York: John Wiley.
- Hox, J. J. (1994), "Hierarchical Regression Models for Interviewer and Respondent Effects," *Sociological Methods & Research*, 22, 300–318.
- Hox, J. J. (2010), *Multilevel Analysis: Techniques and Applications* (2nd ed.), New York: Routledge.
- Loosveldt, G., and K. Beullens (2013a), "The Impact of Respondents and Interviewers on Interview Speed in Face-To-Face Interviews," *Social Science Research*, 42, 1422–1430.
- Loosveldt, G., and K. Beullens (2013b), "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. (2013), "Do Non-Response Follow-Ups Improve or Reduce Data Quality? A Review of the Existing Literature," *Journal of the Royal Statistical Society A*, 176, 129–145.
- Olson, K., and I. Bilgen (2011), "The Role of Interviewer Experience on Acquiescence," *Public Opinion Quarterly*, 75, 99–114.
- Olson, K., and B. Parkhurst (2013), "Collecting Paradata for Measurement Error Evaluations," in *Improving Surveys with Paradata: Analytic Uses of Process Information*, ed. F. Kreuter, pp. 43–72, Hoboken, NJ: John Wiley & Sons.
- Olson, K., and A. Peytchev (2007), "Effect of Interviewer Experience on Interview Pace and Interviewer Attitudes," *Public Opinion Quarterly*, 71, 273–286.
- Olson, K., and J. D. Smyth (2015), "The Effect of CATI Questions, Respondents, and Interviewers on Response Time," *Journal of Survey Statistics and Methodology*, 3, 361–396.
- Olson, K., A. Kirchner, and J. Smyth (2016), "Do Interviewers with High Cooperation Rates Behave Differently? Interviewer Cooperation Rates and Interview Behaviors," *Survey Practice*, 9.
- Ratcliff, R. (1993), "Methods for Dealing with Reaction Time Outliers," *Psychological Bulletin*, 114, 510–532.
- Raudenbush, S. W., and A. S. Bryk (2002), *Hierarchical Linear Models* (2nd ed.), Thousand Oaks, CA: Sage Publications.
- Schaeffer, N. C., J. Dykema, and D. W. Maynard (2010), "Interviewers and Interviewing," in *Handbook of Survey Research* (2nd ed), eds. James D. Wright and Peter V. Marsden, pp. 437–470, UK: Emerald Group Publishing Limited.
- Snijders, T. A. B., and R. J. Bosker (2012), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed). Thousand Oaks, CA, Sage.
- Snijders, G., J. Hox, and E. D. de Leeuw (1999), "Interviewers' Tactics for Fighting Survey Nonresponse," *Journal of Official Statistics*, 15, 185–198.
- Stokes, S. L., and M. Yeh (1988), "Searching for Causes of Interviewer Effects in Telephone Surveys," in *Telephone Survey Methodology*, eds. R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, and J. Waksberg, pp. 357–373. New York: John Wiley and Sons.
- West, B. T., and K. Olson (2010), "How Much of Interviewer Variance is Really Nonresponse Error Variance?" *Public Opinion Quarterly*, 74, 1004–1026.

- Yan, T., and K. Olson (2013), "Analyzing Paradata to Investigate Measurement Error," in *Improving Surveys with Paradata: Analytic Uses of Process Information*, ed. F. Kreuter, pp. 73–96, Hoboken, NJ: John Wiley & Sons.
- Yan, T., and R. Tourangeau (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.

Tables & Appendices follow

Table 1. Distribution of Interviewer and Respondent Characteristics

| Indicator | Operationalization | WLT1 mean before centering (SD) | WLT2 mean before centering (SD) |
|---|---|---------------------------------------|---------------------------------------|
| 1. Experience | | | |
| Interview length | Interview length in minutes excluding respondent selection; top-coded at 1st and the 99th percentile | 12.68 (3.62) | 13.37 (3.28) |
| Interview order | Sequence the interviews were conducted in by a single interviewer | 22.89 (4.70) | 34.58 (15.62) |
| I. experience | Experience within the survey company: 0 years; 1+ years | 73.68% | 30.77% |
| I. cooperation rate | Average interviewer cooperation rate (interview conditional on contact) | 6.84 (1.66) | 11.85 (5.11) |
| 2a. Response propensity: composition | | | |
| R. gender | Gender: 0 = male; 1 = female | 63.91% | 51.72% |
| R. age | Age in years | 61.50 (16.79) | 54.25 (18.38) |
| R. education | Education: 0 = more than high school; 1 = high school degree or less | 28.97% | 31.26% |
| R. race | Race: 0 = white; 1 = nonwhite | 12.64% | 20.47% |
| R. employed | Full-time or part-time employed: 0 = no; 1 = yes | 40.23% | 47.72% |
| R. household income | Total family income (8 categories): 1 = Less than \$10,000; 2 = \$10,000 to \$19,999; 3 = \$20,000 to \$29,999; 4 = \$30,000 to \$39,999; 5 = \$40,000 to \$49,999; 6 = \$50,000 to \$74,999; 7 = \$75,000 to \$99,999; 8 = \$100,000 and above | 4.78 (2.16) | 5.08 (2.41) |

Continued

Table 1. Continued

| Indicator | Operationalization | WLT1 mean before centering (SD) | WLT2 mean before centering (SD) |
|--|--|---------------------------------------|---------------------------------------|
| R. household size | Household size in number of persons: 1 = 1; 2 = 2; 3 = 3; 4 = 4; 5 = 5+ | 2.17 (1.35) | 2.45 (1.43) |
| R. parent | Parent or guardian of children under 18 living in household: 0 = no; 1 = yes | 17.93% | 21.25% |
| R. volunteering | Engaged in any volunteer activities through or for an organization in the last 12 months: 0 = no; 1 = yes | 47.13% | 56.73% |
| 2b. Response propensity: contactability and cooperation | | | |
| R. item non-response rate | Percentage of item nonresponse across the entire survey | 3.05% | 1.75% |
| R. ever refused | Whether a respondent ever refused (soft refusal) to participate in the survey: 0 = no; 1 = yes | 6.70% | 7.80% |
| R. complete at first contact | Was the interview completed at first contact: 0 = no; 1 = yes | 72.28% | 82.98% |
| R. # of call attempts | Number of contact attempts before successful completion | 2.57 (1.58) | 2.21 (1.42) |
| R. time of day when interview completed | Call window when interview was successfully completed: 0 = weekday evening; 1 = weekend evening; 2 = weekend day | 79.31% 9.66% 11.03% | 76.97% 7.45% 15.57% |
| 2c. Response propensity: interaction | | | |
| R. word count | Respondent sum of words spoken during the interview divided by 100 and censored at the 1st and 99th percentile | 5.00 (4.17) | NA |
| I. word count | Interviewer sum of words spoken during the interview divided by 100 and censored at the 1st and 99th percentile | 12.05 (2.53) | NA ^a |

Continued

Table 1. Continued

| Indicator | Operationalization | WLT1 mean before centering (SD) | WLT2 mean before centering (SD) |
|------------------------|--|---------------------------------------|---------------------------------------|
| Controls | | | |
| R. # of questions | Number of questions asked including definition, thank you, and farewell screens; varies due to skip patterns | 50.14 (4.51) | 51.20 (4.40) |
| R. # of answer changes | Number of answers changed once they had been recorded | 0.55 (0.85) | 0.46 (0.74) |
| R. computer usage | Skip indicator: whether respondents triggered follow-up questions related to computer usage | 77.47% | 81.65% |
| I. race | Race: 0 = white; 1 = nonwhite | 52.63% | 46.15% |
| I. gender | Gender: 0 = male; 1 = female | 47.36% | 42.31% |
| I. shift | At least 90.3% of calls made on weekday evening shift. | 15.79% | 26.92% |
| Version | WLT2-specific version indicator 0 = version A 1 = version B | NA | 50.17% |
| Analytic sample size | Respondents Interviewer | n = 435 n = 19 | n = 899 n = 26 |

a. Word count not available in WLT2.

Table 2. Intraclass Correlations for Interview Length, Null Model

| | WLT1 | WLT2 |
|------------------------------------|----------|-----------|
| Interviewer level variance | 2.74** | 2.82*** |
| Respondent level variance | 10.21*** | 8.45*** |
| Likelihood ratio $\chi^2(1)$ | 70.55*** | 153.42*** |
| Intraclass correlation coefficient | 0.21 | 0.25 |

+ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 3. Estimated Coefficients, Standard Errors, and Variance Components for Hierarchical Models Predicting Length of Interview (in Minutes)

| Interview length (trimmed) | WLT1 | WLT 1 (incl. H2c) | WLT2 |
|--|----------|-------------------|----------|
| 1. Experience | | | |
| Ln(interview order) | -0.91* | -0.59** | -0.63*** |
| | (0.36) | (0.19) | (0.12) |
| I. experience (ref. 0 years) | -0.21 | -1.09 | 0.69 |
| | (1.29) | (1.04) | (0.92) |
| I. exp. x ln(int. order) | 0.46 | 0.27 | -0.31 |
| | (0.39) | (0.21) | (0.24) |
| I. cooperation rate | -0.46† | -0.13 | -0.04 |
| | (0.27) | (0.26) | (0.07) |
| 2a. Response propensity: composition | | | |
| R. gender (ref. male) | -0.73* | -0.46** | -0.52** |
| | (0.29) | (0.15) | (0.18) |
| R. age | 0.06*** | 0.02** | 0.02** |
| | (0.01) | (0.01) | (0.01) |
| R. education (ref. college and above) | 0.39 | -0.31† | -0.45* |
| | (0.34) | (0.19) | (0.22) |
| R. race (ref. white) | 1.01*** | 0.04 | 0.42*** |
| | (0.23) | (0.13) | (0.10) |
| R. employed (ref. other) | -2.70*** | -0.22 | 0.94* |
| | (0.69) | (0.38) | (0.49) |
| R. household income | 0.01 | -0.04 | -0.15** |
| | (0.08) | (0.04) | (0.05) |
| R. household size (ref. 1) 2 | -1.10** | -0.45* | 0.09 |
| | (0.35) | (0.19) | (0.25) |
| 3 | -0.84* | -0.20 | 0.50 |
| | (0.49) | (0.26) | (0.33) |
| 4 | -0.88 | -0.50 | 0.95* |
| | (0.65) | (0.35) | (0.39) |
| 5 or more | 0.22 | 0.07 | 1.15** |
| | (0.71) | (0.38) | (0.40) |
| R. parent (ref. no) | 0.08 | 0.07 | -0.52* |
| | (0.51) | (0.27) | (0.27) |
| R. volunteering (ref. no) | 0.07 | 0.14 | -0.04 |
| | (0.31) | (0.16) | (0.19) |
| 2b. Response propensity: contactability and cooperation | | | |
| R. item nonresponse rate | 0.13*** | -0.01 | 0.19*** |
| | (0.04) | (0.02) | (0.03) |
| R. ever refused | -0.40 | -0.17 | -0.33 |
| | (0.62) | (0.33) | (0.44) |
| R. complete at first contact | -0.33 | -0.30 | 0.73* |
| | (0.41) | (0.22) | (0.34) |
| R. # of call attempts | 0.03 | -0.04 | 0.10 |
| | (0.12) | (0.06) | (0.07) |
| R. time of day when interview completed (ref. week eve) | -0.13 | -0.00 | -0.09 |
| | (0.50) | (0.27) | (0.35) |
| weekend & eve | | | |
| weekend & day e | -1.18* | -0.59* | 0.02 |
| | (0.48) | (0.26) | (0.26) |

Continued

Table 3. Continued

| Interview length (trimmed) | WLT1 | WLT 1 (incl. H2c) | WLT2 |
|---|--------------------|--------------------|--------------------|
| 2c. Response propensity: interaction | | | |
| R. word count | | 0.49*** (0.04) | |
| I. word count | | 0.82*** (0.06) | |
| Controls | | | |
| R. # of questions | 0.50*** (0.09) | 0.06 (0.05) | 0.06 (0.06) |
| R. # of answer changes | 0.75*** (0.17) | 0.14 (0.09) | 0.78*** (0.13) |
| R. computer usage | -0.04 (0.43) | 0.61** (0.23) | 0.04 (0.28) |
| I. race (ref. white) | -0.92 (0.78) | -0.68 (0.76) | -0.06 (0.67) |
| I. gender (ref. male) | 1.32 (0.81) | 0.28 (0.79) | 0.97 (0.72) |
| I. shift (ref. not weekday evening) | -1.41 (1.07) | -0.60 (1.04) | -0.24 (0.71) |
| Version | | | 1.15† (0.60) |
| Constant | 15.97*** (1.53) | 15.14*** (1.24) | 13.71*** (0.93) |
| Interviewer-level variance | 2.10† (0.80) | 2.20* (0.76) | 1.98* (0.62) |
| Respondent-level variance | 7.54*** (0.52) | 2.14*** (0.15) | 6.76*** (0.32) |
| Intraclass Correlation Coefficient | 0.22 | 0.51 | 0.23 |
| N | 435 | 435 | 899 |
| AIC | 2,212.80 | 1,690.93 | 4,393.53 |

Standard errors in parentheses.

† p < 0.10 ; * p < 0.05 ; ** p < 0.01 ; *** p < 0.001

APPENDIX**A.1. Proportions and Means in Socio-Demographics for Early, Middle, Late, and Very Late Respondents**

| Indicators of composition | Early interview 1 to 9 | Middle interview 10 to 19 | Late interview 19 to 29 | Very late interview 30 and up | p value: middle vs. early late vs. early very late vs. early (based on LR Chi ² /t-test) |
|--------------------------------------|------------------------------|---------------------------------|-------------------------------|-------------------------------------|--|
| WLT1 | | | | | |
| Female | 0.58 | 0.68 | 0.66 | NA | 0.057 0.266 |
| Age (centered, mean) | 0.78 | -0.84 | 0.73 | NA | 0.372 0.981 |
| High school and less | 0.29 | 0.25 | 0.36 | NA | 0.410 0.297 |
| Nonwhite | 0.14 | 0.20 | 0.31 | NA | 0.343 0.119 |
| Employed | 0.35 | 0.43 | 0.47 | NA | 0.126 0.056 |
| Household income (brackets) | 4.43 | 4.78 | 5.14 | NA | 0.186 0.039 |
| Household size | 2.01 | 2.21 | 2.42 | NA | 0.191 0.013 |
| Parent of children under 18 in HH | 0.13 | 0.20 | 0.23 | NA | 0.072 0.032 |
| Volunteering | 0.42 | 0.49 | 0.54 | NA | 0.172 0.047 |
| Observations WLT1 | 171 | 174 | 90 | 0 | |
| WLT2 | | | | | |
| Female | 0.48 | 0.54 | 0.54 | 0.51 | 0.186 0.270 0.579 |
| Age (centered, mean) | 3.39 | -1.11 | -0.47 | -1.91 | 0.004 0.018 0.001 |
| High school and less | 0.31 | 0.29 | 0.33 | 0.32 | 0.609 0.752 0.766 |
| Nonwhite | 0.27 | 0.41 | 0.43 | 0.46 | 0.080 0.103 0.008 |

Continued

A.1. Continued

| Indicators of composition | Early interview 1 to 9 | Middle interview 10 to 19 | Late interview 19 to 29 | Very late interview 30 and up | p value: middle vs. early late vs. early very late vs. early (based on LR Chi ² /t-test) |
|--------------------------------------|------------------------------|---------------------------------|-------------------------------|-------------------------------------|--|
| Employed | 0.41 | 0.47 | 0.50 | 0.53 | 0.200 0.079 0.016 |
| Household income (brackets) | 4.95 | 5.07 | 5.24 | 5.11 | 0.596 0.214 0.467 |
| Household size | 2.30 | 2.53 | 2.44 | 2.53 | 0.093 0.321 0.087 |
| Parent of children under 18 in HH | 0.15 | 0.21 | 0.23 | 0.26 | 0.105 0.043 0.006 |
| Volunteering | 0.56 | 0.61 | 0.58 | 0.52 | 0.230 0.554 0.458 |
| Observations WLT2 | 237 | 241 | 190 | 234 | |

A.2. Estimated Coefficients, Standard Errors, and Variance Components for Hierarchical Models Predicting Length of Interview (in Minutes)

| Interview length (trimmed) | WLT1 | | | | | WLT2 | | | | | |
|---|------|------------------------------|------------------------------|------------------------------------|-----------------------------|------------------------------|------|--------------------|--------------------|------------------------------------|------------------------------|
| | Null | 1. Experience | 2a. Composition | 2b. Contactability and cooperation | 2c. Interaction (excl. H2c) | Full (excl. H2c) | Null | 1. Experience | 2a. Composition | 2b. Contactability and cooperation | Full |
| 1. Experience | | | | | | | | | | | |
| Ln(interview order) | | -0.63 ⁺ (0.38) | | | | -0.91* (0.36) | | -0.59*** (0.12) | | | -0.63*** (0.12) |
| I. experience (ref. 0 years) | | -0.24 (1.27) | | | | -0.21 (1.29) | | 0.78 (0.94) | | | 0.69 (0.92) |
| I. exp. x Ln(int. order) | | 0.46 (0.43) | | | | 0.46 (0.39) | | -0.34 (0.25) | | | -0.31 (0.24) |
| I. cooperation rate | | -0.51* (0.24) | | | | -0.46 ⁺ (0.27) | | -0.04 (0.07) | | | -0.04 (0.07) |
| 2a. Response propensity: composition | | | | | | | | | | | |
| R. gender (ref. male) | | | -0.78** (0.30) | | | -0.73* (0.29) | | | -0.67*** (0.19) | | -0.52** (0.18) |
| R. age | | | 0.06*** (0.01) | | | 0.06*** (0.01) | | | 0.02** (0.01) | | 0.02** (0.01) |
| R. education (ref. college and above) | | | 0.30 (0.35) | | | 0.39 (0.34) | | | -0.58* (0.23) | | -0.45* (0.22) |
| R. race (ref. white) | | | 0.97*** (0.24) | | | 1.01*** (0.23) | | | 0.35** (0.11) | | 0.42*** (0.10) |
| R. employed (ref. other) | | | -2.61*** (0.70) | | | -2.70*** (0.69) | | | 1.25* (0.51) | | 0.94* (0.49) |
| R. household income | | | -0.03 (0.09) | | | 0.01 (0.08) | | | -0.17*** (0.05) | | -0.15** (0.05) |
| R. household size (ref. 1) 2 | | | -1.08** (0.35) | | | -1.10** (0.35) | | | 0.08 (0.26) | | 0.09 (0.25) |
| 3 | | | -1.01* (0.50) | | | -0.84 ⁺ (0.49) | | | 0.40 (0.34) | | 0.50 (0.33) |
| 4 | | | -1.11 ⁺ (0.67) | | | -0.88 (0.65) | | | 1.09** (0.41) | | 0.95* (0.39) |
| 5 or more | | | -0.09 (0.73) | | | 0.22 (0.71) | | | 1.29** (0.42) | | 1.15** (0.40) |
| R. parent (ref. no) | | | 0.18 (0.52) | | | 0.08 (0.51) | | | -0.73* (0.28) | | -0.52 ⁺ (0.27) |
| R. volunteering (ref. no) | | | 0.09 (0.31) | | | 0.07 (0.31) | | | -0.14 (0.20) | | -0.04 (0.19) |

2b. Response propensity: contactability and cooperation

| | | | | |
|---|-------------------|-------------------|-------------------|-------------------|
| R. item nonresponse rate | 0.16*** (0.04) | 0.13*** (0.04) | 0.20*** (0.03) | 0.19*** (0.03) |
| R. ever refused | -0.58 (0.67) | -0.40 (0.62) | -0.42 (0.47) | -0.33 (0.44) |
| R. complete at first contact | -0.48 (0.44) | -0.33 (0.41) | 0.62+ (0.35) | 0.73* (0.34) |
| R. # of call attempts | -0.09 (0.12) | 0.03 (0.12) | 0.00 (0.07) | 0.10 (0.07) |
| R. time of day when interview completed (ref. week eve) | | | | |
| weekend & eve | 0.20 (0.54) | -0.13 (0.50) | 0.02 (0.37) | -0.09 (0.35) |
| weekend & day | -1.01+ (0.52) | -1.18* (0.48) | 0.22 (0.28) | 0.02 (0.26) |

2c. Response propensity: interaction

| | |
|---------------|-------------------|
| R. word count | 0.52*** (0.04) |
| I. word count | 0.81*** (0.06) |

Controls

| | | | | | | | | | | | |
|-------------------------------------|--------------------|--------------------|--------------------|--------------------|-------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| R. # of questions | 0.09* (0.04) | 0.48*** (0.09) | 0.09* (0.04) | 0.00 (0.02) | 0.50*** (0.09) | 0.11*** (0.02) | 0.01 (0.06) | 0.12*** (0.02) | 0.06 (0.06) | | |
| R. # of answer changes | 0.68*** (0.18) | 0.78*** (0.17) | 0.58*** (0.18) | 0.03 (0.09) | 0.75*** (0.17) | 0.82*** (0.13) | 0.77*** (0.13) | 0.68*** (0.13) | 0.78*** (0.13) | | |
| R. computer usage | -0.31 (0.42) | -0.36 (0.44) | -0.01 (0.42) | 0.47* (0.21) | -0.04 (0.43) | -0.37 (0.26) | -0.09 (0.29) | -0.23 (0.27) | 0.04 (0.28) | | |
| I. race (ref. white) | -1.21+ (0.70) | -1.30+ (0.77) | -1.45+ (0.76) | -0.67 (0.68) | -0.92 (0.78) | 0.16 (0.67) | 0.52 (0.66) | 0.39 (0.67) | -0.06 (0.67) | | |
| I. gender (ref. male) | 1.19* (0.72) | 0.58 (0.76) | 0.69 (0.75) | 0.12 (0.67) | 1.32 (0.81) | 0.92 (0.72) | 0.63 (0.68) | 0.75 (0.68) | 0.97 (0.72) | | |
| I. shift (ref. not weekday evening) | -1.51 (0.94) | -2.05* (1.05) | -1.88+ (1.03) | -0.93 (0.93) | -1.41 (1.07) | -0.23 (0.71) | 0.02 (0.73) | -0.06 (0.73) | -0.24 (0.71) | | |
| Version | | | | | | 0.97 (0.60) | 1.11+ (0.64) | 1.00 (0.64) | 1.15+ (0.60) | | |
| Constant | 12.73*** (0.41) | 13.81*** (1.34) | 15.22*** (0.98) | 13.61*** (0.83) | 0.64 (0.91) | 15.97*** (1.53) | 13.63*** (0.35) | 14.16*** (0.75) | 12.80*** (0.80) | 11.97*** (0.69) | 13.71*** (0.93) |
| Interviewer-level variance | 2.74** (1.03) | 1.52 (0.63) | 2.31* (0.87) | 2.17* (0.84) | 2.00* (0.69) | 2.10+ (0.80) | 2.82*** (0.87) | 1.99* (0.63) | 2.28** (0.71) | 2.33** (0.72) | 1.98* (0.62) |
| Respondent-level variance | 10.21*** (0.71) | 9.67*** (0.67) | 8.04*** (0.56) | 9.27*** (0.64) | 2.43*** (0.17) | 7.54*** (0.52) | 8.45*** (0.40) | 7.63*** (0.37) | 7.43*** (0.36) | 7.54*** (0.36) | 6.76*** (0.32) |
| N | 435 | 435 | 435 | 435 | 435 | 435 | 899 | 899 | 899 | 899 | 899 |
| AIC | 2,288.04 | 2,276.03 | 2,221.57 | 2,267.73 | 1,698.21 | 2,212.80 | 4,539.06 | 4,464.38 | 4,459.26 | 4,461.61 | 4,393.53 |

Standard errors in parentheses;

† p < 0.10 ; * p < 0.05 ; ** p < 0.01 ; *** p < 0.001

A.3. Estimated Coefficients (Excluding the Interaction Effect), Standard Errors, and Variance Components for Hierarchical Models Predicting Length of Interview (in Minutes)

| Interview length (trimmed) | WLT1 | WLT 1 (incl. H2c) | WLT2 |
|--|--------------------|--------------------|--------------------|
| 1. Experience | | | |
| Ln(interview order) | -0.55** (0.19) | -0.38*** (0.10) | -0.70*** (0.11) |
| I. experience (ref. 0 years) | 0.79 (0.97) | -0.52 (0.94) | -0.07 (0.72) |
| I. cooperation rate | -0.46* (0.27) | -0.13 (0.26) | -0.04 (0.07) |
| 2a. Response propensity: composition | | | |
| R. gender (ref. male) | -0.74* (0.29) | -0.47** (0.15) | -0.51** (0.18) |
| R. age | 0.06*** (0.01) | 0.02** (0.01) | 0.02** (0.01) |
| R. education (ref. college and above) | 0.37 (0.34) | -0.32 (0.19) | -0.44* (0.22) |
| R. race (ref. white) | 1.02*** (0.23) | 0.05 (0.13) | 0.41*** (0.10) |
| R. employed (ref. other) | -2.66*** (0.69) | -0.20 (0.38) | 0.97* (0.49) |
| R. household income | -0.00 (0.08) | -0.05 (0.04) | -0.15** (0.05) |
| R. household size (ref. 1) 2 | -1.10** (0.35) | -0.45* (0.19) | 0.10 (0.25) |
| 3 | -0.84+ (0.49) | -0.20 (0.26) | 0.50 (0.33) |
| 4 | -0.90 (0.65) | -0.52 (0.35) | 1.00* (0.39) |
| 5 or more | 0.19 (0.71) | 0.05 (0.38) | 1.17** (0.40) |
| R. parent (ref. no) | 0.14 (0.50) | 0.10 (0.27) | -0.54* (0.27) |
| R. volunteering (ref. no) | 0.08 (0.31) | 0.15 (0.16) | -0.03 (0.19) |
| 2b. Response propensity: contactability and cooperation | | | |
| R. item nonresponse rate | 0.12* (0.04) | -0.01 (0.02) | 0.19*** (0.03) |
| R. ever refused | -0.34 (0.62) | -0.14 (0.33) | -0.33 (0.44) |
| R. complete at first call | -0.31 (0.41) | -0.30 (0.22) | 0.72* (0.34) |
| R. # of call attempts | -0.02 (0.12) | -0.05 (0.06) | 0.09 (0.07) |

*Continued***A.3.** Continued

| Interview length (trimmed) | WLT1 | WLT 1 (incl. H2c) | WLT2 |
|---|--------------------|--------------------|--------------------|
| R. call window (ref. week eve) wend eve | -0.07 (0.50) | 0.04 (0.26) | -0.09 (0.35) |
| wend day | -1.15* (0.48) | -0.57* (0.26) | 0.01 (0.26) |
| 2c. Response propensity: interaction | | | |
| R. word count | | 0.49*** (0.04) | |
| I. word count | | 0.81** (0.06) | |
| Controls | | | |
| R. # of questions | 0.50*** (0.09) | 0.06 (0.05) | 0.06 (0.06) |
| R. # of answer changes | 0.75*** (0.17) | 0.14 (0.09) | 0.77*** (0.13) |
| R. computer usage | -0.08 (0.43) | 0.59* (0.23) | 0.04 (0.28) |
| I. race (ref. white) | -0.90 (0.78) | -0.67 (0.75) | -0.07 (0.68) |
| I. gender (ref. male) | 1.30 (0.81) | 0.27 (0.78) | 0.96 (0.73) |
| I. shift (ref. not weekday evening) | -1.39 (1.07) | -0.59 (1.03) | -0.24 (0.71) |
| Version | | | 1.14+ (0.61) |
| Constant | 15.25*** (1.41) | 14.72*** (1.20) | 13.90*** (0.92) |
| Interviewer-level variance | 2.11+ (0.80) | 2.19* (0.75) | 2.02* (0.63) |
| Respondent-level variance | 7.56*** (0.52) | 2.14*** (0.15) | 6.77*** (0.32) |
| N | 435 | 435 | 899 |
| AIC | 2,212.18 | 1,690.54 | 4,393.24 |

Standard errors in parentheses;

† p < 0.10 ; * p < 0.05 ; ** p < 0.01 ; *** p < 0.001