

First Impressions of Telephone Survey Interviewers

Jessica Broome¹

Survey nonresponse may increase the chances of nonresponse error, and different interviewers contribute differentially to nonresponse. This article first addresses the relationship between initial impressions of interviewers in survey introductions and the outcome of these introductions, and then contrasts this relationship with current viewpoints and practices in telephone interviewing. The first study described here exposed judges to excerpts of interviewer speech from actual survey introductions and asked them to rate twelve characteristics of the interviewer. Impressions of positive traits such as friendliness and confidence had no association with the actual outcome of the call, while higher ratings of “scriptedness” predicted lower participation likelihood. At the same time, a second study among individuals responsible for training telephone interviewers found that when training interviewers, sounding natural or unscripted during a survey introduction is not emphasized. This article concludes with recommendations for practice and further research.

Key words: Survey; telephone; nonresponse; interviewer.

1. Introduction and Background

Survey nonresponse has the potential to bias survey estimates (Groves et al. 2004). It has been demonstrated that telephone interviewers vary substantially in their response rates (Oksenberg and Cannell 1988). Identifying vocal characteristics and techniques of more successful telephone interviewers (i.e., those with higher overall response rates) may impact data quality by allowing for more targeted screening and training of interviewers with the aim of reducing nonresponse.

Literature from both survey methodology (Oksenberg et al. 1986) and telemarketing (Ketrow 1990) has found that a pleasing or attractive voice in the initial seconds of a phone call is imperative in extending the interaction. Further, Ketrow (1990) discusses the importance of giving an initial impression of competence, and Oksenberg and colleagues (Oksenberg et al. 1986; Oksenberg and Cannell 1988) found that judges’ ratings of phone-interviewer competence based on brief recorded excerpts were positively associated with the interviewers’ success. This is not to imply that in survey interview introductions, having a pleasing, competent-sounding voice in the opening statement is enough to guarantee success. However, an interviewer voice that gives listeners a positive first

¹ University of Michigan – Institute for Social Research, 426 Thompson Street, Ann Arbor, Michigan 48104, U.S.A. Email: jsbroome@umich.edu

Acknowledgments: The author would like to acknowledge the Charles Cannell Fund in Survey Methodology, the University of Michigan Center for the Education of Women, the A. Regula Herzog Young Investigators Fund, and the Harvey G. and Joyce H. Behner Graduate Fellowship for their support as well as Frederick Conrad, Norbert Schwarz, Steve Heeringa, and Susan Brennan for their input.

impression may lead to a longer conversation, thus increasing the likelihood of participation.

Nonresponse to telephone surveys has been increasing steadily over the past 25 years (Curtin et al. 2005). Declining response rates have the potential to increase nonresponse error (Groves et al. 2004; Teitler et al. 2003). Further, nonresponse rates vary by interviewer (Morton-Williams 1993; Oksenberg and Cannell 1988; O'Muircheartaigh and Campanelli 1999; Snijders et al. 1999). Uncovering the characteristics and tactics of successful interviewers can help to reduce nonresponse, either by using vocal and personality characteristics as hiring criteria or by training interviewers to adopt characteristics or tactics which have been shown to lead to increased success.

In contrast to face-to-face interviewers, telephone survey interviewers have just two primary tools that are under their control in their efforts to persuade answerers to participate: what they say (speech) and how they say it (vocal characteristics). A small body of literature (e.g., Sharf and Lehman 1984; Oksenberg et al. 1986; Oksenberg and Cannell 1988; Groves et al. 2007; Conrad et al. 2013) finds relationships between vocal characteristics of interviewers in telephone-survey introductions and interviewer success in obtaining interviews. In general, successful interviewers have been ones who spoke louder (Oksenberg et al. 1986; Oksenberg and Cannell 1988; van der Vaart et al. 2005) and with more falling intonation (Sharf and Lehman 1984; Oksenberg and Cannell 1988). In addition, success has been shown to be correlated both with higher mean fundamental frequency (Sharf and Lehman 1984) and higher perceived pitch (Oksenberg et al. 1986), as well as variable fundamental frequency (Sharf and Lehman 1984; Groves et al. 2007) and variable pitch (Oksenberg et al. 1986). The terms "pitch" and "fundamental frequency" are often used interchangeably, but a necessary distinction is that fundamental frequency is an acoustic measure of vocal-chord vibrations, while pitch is a listener's perception of frequency or how "high" or "low" a voice sounds.

Three recent studies have found nonlinear relationships between success and rate of speech (Groves et al. 2007; Steinkopf et al. 2010; Benkí, Broome, Conrad, Groves and Kreuter 2011): contacts with speech that is either overly slow or overly fast tend to be less successful. Benkí et al. (2011) found that contacts with interviewer speech in the range of 3.34–3.68 words per second were the most likely to be successful.

One critical question concerns what underlies these associations; what is it about an interviewer who speaks at a particular rate or with more variable pitch that leads to success, especially given the limited amount of exposure an answerer has to the interviewer's voice before deciding whether or not to participate? Oksenberg et al. (1986, 99) emphasized the importance for an interviewer to have a voice that potential respondents find appealing in the first few seconds of a survey interview introduction context, stating that "if vocal characteristics lead the respondent to perceive the interviewer as unappealing, cooperation will be less likely."

Two dimensions of person perception, warmth and competence, have been shown to be relevant to the development of first impressions of others across a range of contexts (Asch 1946; Fiske, Cuddy, and Glick 2007; Kelley 1950; Rosenberg, Nelson, and Vivekanathan 1968). Several studies in the literature on interviewer vocal characteristics (Oksenberg et al. 1986; van der Vaart et al. 2005) suggest that ratings of personal characteristics on these dimensions of person perception are associated with both interviewer response rates

and vocal characteristics. These studies involved collecting ratings of several interviewer personality characteristics, which were then successfully reduced to two dimensions interpretable as “warmth” and “competence.” Characteristics on the “warmth” dimension included being cheerful, friendly, enthusiastic, polite, interested in her task, and pleasant to listen to. [Oksenberg et al. \(1986\)](#) and [van der Vaart et al. \(2005\)](#) found correlations between high ratings on the warmth dimension and vocal characteristics, including variation in pitch, higher pitch, and a faster rate of speech, suggesting that listeners’ impressions of interviewer personality are based, at least in part, on physical (acoustic) attributes of interviewers’ voices. Characteristics composing the “competence” dimension included being self-assured, educated, intelligent, and professional. [Van der Vaart et al. \(2005\)](#) found that interviewers rated highly on “competence” characteristics tended to have lower pitch.

Importantly, [Oksenberg et al. \(1986\)](#) and [Van der Vaart et al. \(2005\)](#) found that high ratings on a “warmth” dimension correlated with ratings of judges’ willingness to participate. This aligns with [Morton-Williams’s \(1993\)](#) finding that warm or “likable” interviewers increased perceived benefits to potential respondents and improved participation rates, and also with [Cialdini’s \(1984\)](#) “Liking” Principle of Compliance: people are more likely to comply with a request from someone they like.

Further, [Cialdini \(1984\)](#) suggests a compliance heuristic based on the principle of authority; requests from an authoritative speaker are more likely to be honored than requests with less authority. Impressions of authoritative characteristics such as competence and confidence, in turn, have been shown to be associated with interviewer success ([Oksenberg et al. 1986](#); [Oksenberg and Cannell 1988](#); [Steinkopf et al. 2010](#)).

While a small body of literature explores the relationship between interviewer vocal characteristics and impressions, there are clearly challenges to conducting research in this area. For example, the independent variables used are judges’ ratings of an interviewer’s vocal characteristics. When such ratings are collected in person, small sample sizes tend to be the norm; for example, two early studies ([Oksenberg et al. 1986](#); [Oksenberg and Cannell 1988](#)) were each based on six recordings. Studies with larger numbers of judges, such as those by [Huefken and Schaefer \(2003\)](#), with 51 judges, [Steinkopf et al. \(2010\)](#), which used 56 judges, and [Van der Vaart et al. \(2005\)](#), with twelve judges, were based on the work of student (rather than professional) interviewers, limiting the applicability of findings. Finally, dependent variables assessed in existing studies are either interviewers’ historical response rates, judges’ own willingness to comply, or judges’ beliefs that “someone” will comply; no study has yet associated vocal characteristics with actual contact outcomes.

The aim of the exploratory studies described in this article was to see whether first impressions, formed in the initial seconds of a telephone interviewer’s introduction, are an important component in determining the outcome of a survey introduction. This article will address several questions concerning first impressions of telephone interviewers:

- Which first impressions are most predictive of a successful outcome?
- How do relationships between first impressions and success compare with practitioners’ ideas about what makes a successful interviewer?

The first hypothesis addressed in this article (h1) is that interviewers who are perceived more positively and less negatively in the initial seconds of a contact by judges will have greater success, as measured by contact outcome.

The second hypothesis (h2) is that the characteristics that practitioners perceive as important to interviewer success will parallel those characteristics that predict actual cooperation.

This article reports results from two studies. The “Listeners’ Study” elicited ratings of interviewer personality characteristics in audio-recorded telephone introductions from five surveys. The raters (or “listeners”) were internet-survey panel members who answered questions after listening to brief excerpts of interviewer speech from real (not staged) telephone-survey introductions. Having a large number of raters, combined with the use of real contacts conducted by professional interviewers for which the actual outcome is known, is unique in studies on interviewer voice. Another novel element of this study is that, in order to explore whether practitioners focus on the attributes of interviewer speech that are most related to the outcome of survey invitations, results from the Listeners’ Study are contrasted with those from a web survey of practitioners who hire and train telephone interviewers. In this “Practitioners’ Study,” practitioners were asked which characteristics of interviewers they consider in hiring and training.

This article concludes with a discussion of implications for practice, as well as suggestions for future research in this area.

2. Data and Methods

The data described in this section are drawn from two web surveys. The “Listeners’ Study” was a survey among 3,403 adult, English-speaking members of an internet-survey panel. The “Practitioners’ Study” was a smaller survey of 44 survey practitioners who are responsible for the hiring and training of survey interviewers in academic, government, and for-profit survey organizations.

2.1. *Listeners’ Study: Selection of Contacts*

The recordings used in the listeners’ study were selected from 1,380 audio-recorded telephone-survey introductions. These introductions, conducted by 100 interviewers, were from five telephone surveys that were audio recorded for another project. In this project, all contacts associated with selected households, regardless of who the interviewer was, were included in the dataset (Benkí et al. 2011; Conrad et al. 2013). Contacts by 49 different interviewers with ranging lengths of tenure and response rates varying over the course of their tenure at University of Michigan from .07–.21 are included in this dataset.

The recordings were classified into five outcomes: “agree,” where the answerer cooperates and agrees to participate; “refuse,” where there is an explicit refusal (for example, “I will not take the survey. Please do not call again”); “scheduled callback,” where the interviewer either schedules a time to call back or asserts that she will call again; “hang up,” where the answerer hangs up but never clearly refuses; and “other.”

The listeners’ study uses excerpts from these recorded introductions (referred to hereafter as “contacts”) from three of the five studies. To facilitate comparisons (particularly in analyses of vocal characteristics such as pitch), only introductions by

female interviewers were selected. Contacts in which the answerer hangs up during or directly following the interviewer's first speaking turn were excluded, using the rationale that these are "hard-core nonrespondents" who are determined not to become respondents, and the interviewer has no opportunity to use her voice or speech to convince them otherwise.

Because listeners were asked to make judgments about the interviewer's personality, contacts had to contain enough speech to make these determinations. The minimum amount of speech required for inclusion was a statement of name and affiliation. Finally, contacts were omitted if the interviewer asked for a particular person by name, indicating that the interviewer had already spoken at length to someone in the household, and the persuasion process was likely to be quite different.

Applying these criteria to the 1,380 contacts resulted in 283 recordings from the Survey of Consumer Attitudes, or SCA ($n = 168$); the National Study on Medical Decisions, or NSMD ($n = 110$); and the Mississippi Community Study, or MCS ($n = 5$).

These 283 contacts form the basis of the listeners' study. 118 (42 percent) had an outcome of "agree" and 165 (58 percent) had an outcome of "refuse." Listeners were not told the likelihood of either outcome.

2.2. *Listeners' Study: Description of Stimulus and Lines of Questioning*

The listeners' study used online presentation of audio recordings of telephone-survey invitations to elicit listeners' judgments about telephone interviewers' personality characteristics. In this study, 3,403 members of an online survey panel listened to interviewer speech from the contacts described above. The stimuli to which listeners were exposed consisted of brief introductory statements by the interviewer, such as: "Hello, my name is _____ and I'm calling from the University of Michigan about our survey on _____."

Each listener heard excerpts from five contacts randomly selected from the corpus described above, which contained 283 introductions by 49 different interviewers. It was possible for some listeners to hear multiple introductions by one interviewer, and for others to hear five different interviewers. Interviewers had between 1 and 23 contacts in the dataset, with an average of 5.8 contacts. While the same group of five contacts could conceivably be heard by multiple listeners, assignment and order of excerpts were random so as to avoid context effects from presenting excerpts in set groups or a set order.

For each of the five contacts, listeners were asked to rate the interviewer on twelve characteristics (confident, competent, professional, knowledgeable, enthusiastic, pleasant to listen to, friendly, natural sounding, genuine, scripted, irritating, and uncertain), using a scale from 1 (not at all) to 6 (extremely). These are referred to as "characteristic ratings" below. Many of these traits have been shown to be related to interviewer success in the literature (Oksenberg and Cannell 1988; van der Vaart et al. 2005).

2.3. *Listeners' Study: Preparation of Contacts*

After selecting the interviewer speech to be used, the recording was amplified to use the full range of sounds that a recorded voice would make. Amplification was maintained at the same level for all contacts, thus making all contacts comparable in volume. Finally, to

preserve interviewers' anonymity, the interviewer's name in each contact was replaced with a quarter-second-long tone. For consistency, this was done even in the few cases where the interviewer only said her first name.

2.4. Listeners' Study: Data Collection and Respondent Descriptives

Data collection was conducted by a commercial vendor, Lightspeed Research (<http://www.lightspeedresearch.com/>). 15,000 invitations were sent to a stratified random sample of members of Lightspeed's own 1.3 million-member volunteer online panel. (All sample members had a known chance of being invited; the list of invitees was stratified by gender, age and region in an attempt to attain representativeness of the US population on these variables.) This panel is recruited using a variety of sources, including opt-in email, co-registration, e-newsletter campaigns, and placements of banner advertisements. Panelists receive regular survey invitations. While the panel does not perfectly mirror the gender distribution of the US population according to Census data (32% of panelists are male, compared to 49% in the population), the respondents to the listeners' survey were more representative in terms of gender. Respondents were evenly divided between males (49%) and females (51%). One-third (33%) were aged 60 or older, while 20% were 50–59, 18% were 40–49, 17% were 30–39, and 12% were 18–29. 88% of respondents were white (compared to 68% in the general population), and 81% had at least some college education (compared to 55% of the national population, according to the 2010 US Census). These discrepancies can be considered a limitation of the study.

This study was fielded August 12–18, 2011. Panel members were screened to ensure that they were 18 years of age or older (as would be any eligible respondents to the surveys represented by these contacts), and that they characterized their ability to understand spoken English as “excellent” or “good.” This screening criterion was deemed necessary for listeners to be expected to make personality judgments about the interviewer based on brief speech clips.

After their eligibility for the study was determined, listeners were exposed to an “introductory” audio clip and asked to identify one of the words in the clip. The purpose of this exercise was threefold: first, to ensure that listeners were using a computer with working audio; second, to familiarize them with the type of audio they would be hearing during the survey; and third, as a quality-control check to ensure that listeners could sufficiently distinguish words in the contact. 126 potential listeners were screened out at this stage; 3,403 listeners completed the survey.

While the mean exposure length of contacts was 10.32 seconds, the range was wide: from 2.3 to 49.2 seconds. To roughly match the burden on listeners and ensure that none received multiple long contacts, contacts were stratified into five groups based on logical length categories. There were between 45 and 70 contacts in each length category, resulting in more ratings for some contacts than others; the mean number of ratings by length category ranged from 49 to 76. Each listener was exposed to a set of five introductions, each consisting of one randomly selected contact from each length category. For each contact, listeners rated the interviewer on the twelve characteristics discussed above.

2.5. Practitioners' Study

As mentioned above, a second study was conducted among individuals responsible for hiring and training telephone interviewers. A questionnaire was created to assess these practitioners' ratings of the importance of various behaviors and attributes to telephone interviewers' success, as well as to understand their current focuses in hiring and training telephone interviewers. The final survey was programmed in the online survey tool Qualtrics.

A sampling frame was developed that relied on personal contacts of the author, as well as on a list of all members of the Association of Academic Survey Research Organizations (AASRO), which, although not a complete listing of academic survey institutions, was fairly comprehensive and readily accessible. The final sample consisted of 113 individuals at 108 organizations, including two government, 92 academic, three non-profit and eleven for-profit organizations. Two weeks after the initial invitation was sent, a reminder email was sent to all members of the original frame with working email addresses, with the exception of those participants who had already provided their email addresses (respondents were given the option to provide their email addresses if they wished to receive a copy of the results), and those sample members who had requested no further contact.

The survey was completed by 44 respondents between June 5 and July 12, 2011, resulting in a 42% response rate. This response rate is sufficient for the purposes of this study, which was not to uncover precise estimates but rather general trends among practitioners. Further, variation in the organizational characteristics (number of CATI stations and number of interviews conducted in 2010) of those who did respond reduces the chances of nonresponse bias.

Respondents represented a wide range of organizations in size and workload. The median number of CATI stations in respondent organizations was 25, but the number of stations ranged from nine to 450. Close to half (42%) of respondents reported that their organization had conducted fewer than 5,000 telephone interviews in 2010, while an equal percentage reported that their organization had conducted 10,000 or more interviews. Respondents reported that, on average, 81% of the interviews their organizations conducted were for government, academic, or non-profit organizations, 15% were for for-profit organizations, and 2% were for "other" organizations (2% were not sure).

To qualify for the study, practitioners had to have responsibility for hiring and/or training telephone interviewers. Of the 44 respondents, 41 indicated responsibility for hiring interviewers and 40 indicated responsibility for training interviewers.

3. Results

3.1. Listeners' Study: Characteristic Ratings

3.1.1. Description of Ratings

The listeners' study asked for ratings of twelve characteristics of interviewers from five contacts per listener on a six-point scale (1 = not at all to 6 = extremely). Each contact was rated by at least 30 listeners. Analyses were conducted at the contact level; for each

contact, a mean rating across all listeners who heard it was calculated for each characteristic. The mean of all contact-level means, as well as the minimum and maximum mean for each characteristic, are reported. The mean ratings for each characteristic across contacts ranged from 2.50 to 3.89, as shown in [Table 1](#).

Ratings for all nine positive characteristics were highly correlated, as shown in [Table 2](#), indicating that an overall impression of positivity drives judgments.

In addition, a factor analysis, using ratings of all characteristics except scriptedness, revealed that only one factor, explaining 86% of total variance, had extremely high loadings for all nine positive characteristics. The method of principal factors was used to extract factor scores. The overall Kaiser-Meyer-Olkin measure of sampling adequacy, which gives a measure of how much each item is correlated with the others, was 0.92.

Ratings of “uncertain” and “irritating” were highly correlated with each other (.71), but ratings of “scripted” were not highly correlated with ratings of any other characteristic. The mean correlation between scripted and positive characteristics was .01.

3.1.2. Characteristic Ratings As Predictors of Contact Outcome

It was hypothesized (h1) that when ratings of nine positive interviewer characteristics (enthusiastic, friendly, natural, genuine, pleasant to listen to, confident, professional, competent, and knowledgeable) were high and ratings of three negative characteristics (irritating, uncertain, and scripted) were low, a contact would be more likely to result in cooperation than when the positive characteristics were rated lower and the negative characteristics were rated higher.

This hypothesis was partially supported. Twelve bivariate logistic models were constructed, all accounting for the multilevel structure of this dataset (contacts nested within interviewers). For all of these models, the dependent variable was coded as $y = 1$ (agree) or $y = 0$ (refusal). The equation for these models can be written as $\log(\pi_{\text{agree}}/1 - \pi_{\text{agree}}) = a + b_1x_1 + u_j$, where π_{agree} denotes the probability of cooperation; a is an intercept; x_1 represents the mean rating of a characteristic

Table 1. Description of characteristic ratings

| Characteristic | Mean of contact-level means (sd) | Minimum contact- level mean rating | Maximum contact-level mean rating | Spread |
|-----------------------|--|--|---|--------|
| Confident | 3.62 (.56) | 1.77 | 4.63 | 2.86 |
| Professional | 3.70 (.52) | 1.85 | 4.73 | 2.88 |
| Pleasant to listen to | 3.54 (.46) | 2.13 | 4.49 | 2.36 |
| Competent | 3.67 (.50) | 1.90 | 4.67 | 2.77 |
| Knowledgeable | 3.61 (.49) | 2.13 | 4.75 | 2.62 |
| Natural sounding | 3.65 (.41) | 2.35 | 4.47 | 2.12 |
| Enthusiastic | 3.43 (.50) | 2.25 | 4.51 | 2.26 |
| Genuine | 3.59 (.39) | 2.41 | 4.45 | 2.04 |
| Scripted | 3.78 (.31) | 2.70 | 4.67 | 1.97 |
| Friendly | 3.89 (.38) | 2.84 | 4.62 | 1.78 |
| Uncertain | 2.70 (.50) | 1.83 | 4.55 | 2.72 |
| Irritating | 2.50 (.35) | 1.73 | 3.51 | 1.78 |

Table 2. Correlations between characteristic ratings

| | CON | FRI | PRO | PLE | COM | KNO | NAT | ENT | GEN | SCR | UNC | IRR |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|
| Confident | 1.00 | | | | | | | | | | | |
| Friendly | 0.77 | 1.00 | | | | | | | | | | |
| Professional | 0.94 | 0.72 | 1.00 | | | | | | | | | |
| Pleasant | 0.86 | 0.86 | 0.88 | 1.00 | | | | | | | | |
| Competent | 0.97 | 0.76 | 0.97 | 0.88 | 1.00 | | | | | | | |
| Knowledgeable | 0.92 | 0.70 | 0.92 | 0.82 | 0.94 | 1.00 | | | | | | |
| Natural | 0.84 | 0.81 | 0.88 | 0.93 | 0.88 | 0.83 | 1.00 | | | | | |
| Enthusiastic | 0.78 | 0.88 | 0.66 | 0.72 | 0.73 | 0.68 | 0.69 | 1.00 | | | | |
| Genuine | 0.87 | 0.84 | 0.88 | 0.91 | 0.90 | 0.89 | 0.93 | 0.75 | 1.00 | | | |
| Scripted | 0.13 | -0.13 | 0.16 | 0.01 | 0.11 | 0.13 | -0.11 | -0.10 | -0.09 | 1.00 | | |
| Uncertain | -0.89 | -0.63 | -0.85 | -0.72 | -0.85 | -0.77 | -0.69 | -0.65 | -0.70 | -0.17 | 1.00 | |
| Irritating | -0.70 | -0.63 | -0.76 | -0.84 | -0.76 | -0.66 | -0.78 | -0.49 | -0.73 | -0.01 | 0.71 | 1.00 |

(for example, scriptedness) across all listeners who rated the contact; and u_j represents the random, unobserved effects of interviewers.

Of these twelve models, only the model for scripted had a significant coefficient ($b = -1.05$, standard error = .40, $z = -2.59$, $p = 0.010$), indicating that perceptions of the interviewer as more scripted decrease the likelihood of a contact’s success. These results persisted when the models controlled for the length of exposure, and also when only the contacts with the longest exposure lengths (at least ten seconds) were analyzed.

A model was constructed which predicted contact outcome using the factor score described above and the contact’s mean scriptedness rating, while controlling for recording length and accounting for clustering by interviewer. The equation for this model was $\log(\pi_{\text{agree}}/1 - \pi_{\text{agree}}) = a + b_1x_1 + b_2x_2 + b_3x_3 + u_j$, where π_{agree} denotes the probability of cooperation; a is an intercept; x_1 represents the contact’s factor score; x_2 represents the mean rating of scriptedness across all listeners who rated the contact; x_3 represents the length of the recording; and u_j represents the random, unobserved effects of interviewers. As Table 3 shows, only scriptedness was a significant predictor in this model ($z = -2.65$, $p = 0.008$); the factor score was not, indicating that initial impressions of scriptedness, but not of any other characteristic, are important to a contact’s outcome.

In summary, there was no support for the hypothesis that positive characteristics would predict a successful outcome. But a negative characteristic, scriptedness, was negatively associated with success, with contacts where interviewers are less scripted being more successful than those who were rated as more scripted. Agreement with the survey request exhibits almost no variation across interviewers after accounting for other factors including scriptedness, but substantial variation across contacts; in contacts where the interviewer is not considered scripted, agreement is more likely.

3.2. Comparison Between Listeners’ and Practitioners’ Surveys

Hypothesis 2 was that practitioners’ views of what makes a successful interviewer would align with the characteristics which were found to predict contact success in the listeners’ study. To some degree, this is the case; practitioners recognize “how genuine the interviewer sounds” and “the ability to respond to concerns expressed by potential respondents” as important to an interviewer’s response rate, acknowledging that interviewers should not sound robotic during their introductions. However, practitioners appear conflicted; while they recognize the need for a genuine-sounding interaction between interviewer and answerer, they also emphasize the need for interviewers to follow a script, with 48% saying “an interviewer’s ability to follow a script during an

Table 3. Coefficients in model predicting cooperation

| | Coefficient | SE | Z | P |
|---------------------|-------------|-----|--------|-------|
| Factor score | − .017 | .12 | − 0.14 | 0.89 |
| Scriptedness rating | − 1.07 | .40 | − 2.65 | 0.008 |
| Length of recording | .016 | .02 | .98 | 0.33 |

Standard deviation of random interviewer effects: .0000227

Variance of random interviewer effects: 5.0625E-10

Intraclass correlation coefficient: 1.53881E-10

introduction” is extremely important, and 38% rating it as somewhat important. Practitioners rank “an interviewer’s ability to ‘ad lib’ or deviate from a script during an introduction” as slightly less important to an interviewers’ success. On the other hand, results from the listeners’ study indicated that impressions of scriptedness are, in fact, detrimental to the success of contacts, with lower ratings of scriptedness found in successful contacts; indeed, scriptedness is the *only* characteristic that matters to listeners.

Further, of the 18 elements tested (shown in Table 4), the one judged by survey practitioners as most important to an interviewer’s success in obtaining interviews was “the initial impression an interviewer gives to sample members.” This “initial impression” may well include an interviewer’s degree of scriptedness; however, this finding shows that emphases differ between practitioners and listeners, as findings from the listeners’ survey indicate that, aside from scriptedness, no ratings of interviewer characteristics based on early impressions can predict success on a given contact.

Far more important than an ability to “ad lib”, according to practitioners, were traits such as competence, professional demeanor, and confidence – ratings of which were not predictive of actual contact-level outcome.

Among practitioners responsible for training telephone interviewers, just 15% reported that “developing a personalized or nonscripted introduction” is a primary focus of their organization’s interviewer training, while 44% reported that it is not a focus at all. (Practitioners were shown a list of 13 items and given the instruction, “For each of the following, please indicate if it is a primary focus, a secondary focus, or not a focus at all in telephone-interviewer training.”) “Following introductory scripts,” by contrast, was a primary training focus in the vast majority (78%) of organizations surveyed. This emphasis on following introductory scripts contrasts with the finding in the listeners’ study that higher ratings of scriptedness predict less success at the contact level.

These results demonstrate a disconnect between listeners and practitioners. While listeners’ judgments of scriptedness are predictive of a contact’s success (indeed, this is the only rated characteristic associated with success), practitioners place less emphasis on reducing scriptedness and more on other impressions conveyed by interviewers.

4. Conclusions, Applications, and Discussion

This exploratory research has found that survey practitioners believe that initial impressions of an interviewer are important to that interviewer’s success. By contrast, most ratings of interviewer traits such as competence, confidence, and professionalism based on a brief exposure are not predictive of the ultimate outcome of the conversation. One exception to this is ratings of scriptedness, which are significant predictors of contact outcome; indeed, scriptedness is the only rated component of a first impression to predict success. This may be attributable to scriptedness being the most noticeable of all characteristics rated, overwhelming characteristics such as friendly and professional in the initial seconds of an introduction. However, practitioners emphasize the importance of “following a script,” even though this practice might actually harm interviewers’ chances of obtaining a completed interview.

This can be applied to survey practice, as an emphasis on decreasing the scripted or unnatural nature of survey introductions may well serve to improve interviewer

Table 4. Practitioners' ratings of importance to interviewer's success

| | Scale: 1 (not at all important) – 4 (very important) | Mean (sd) | % Extremely important | % Somewhat/ extremely important | % Not very/ not at all important |
|----|--|------------|-----------------------|---------------------------------|----------------------------------|
| 1 | The initial impression an interviewer gives to sample members. | 3.88 (.33) | 88% | 100% | 0% |
| 2 | The ability to address concerns expressed by potential respondents. | 3.84 (.38) | 83% | 100% | 0% |
| 3 | How competent the interviewer sounds to potential respondents. | 3.84 (.38) | 83% | 100% | 0% |
| 4 | Professional demeanor when talking to potential respondents. | 3.81 (.44) | 86% | 98% | 2% |
| 5 | How confident the interviewer sounds to potential respondents. | 3.81 (.45) | 83% | 98% | 2% |
| 6 | The ability to convey knowledge about the study. | 3.72 (.46) | 71% | 100% | 0% |
| 7 | How genuine the interviewer sounds to potential respondents. | 3.70 (.47) | 69% | 100% | 0% |
| 8 | An interviewer's voice that does not sound monotonous (has pitch variability). | 3.58 (.5) | 57% | 100% | 0% |
| 9 | How friendly the interviewer sounds to potential respondents. | 3.53 (.55) | 57% | 98% | 2% |
| 10 | The interviewer's speech rate. | 3.42 (.59) | 48% | 95% | 5% |
| 11 | How enthusiastic the interviewer sounds to potential respondents. | 3.42 (.63) | 50% | 93% | 7% |
| 12 | A pleasant-sounding voice. | 3.41 (.59) | 45% | 95% | 5% |
| 13 | The interviewer's ability to follow a script during an introduction. | 3.30 (.78) | 48% | 86% | 14% |
| 14 | The interviewer speaks without any "um's" or "uh's." | 3.16 (.71) | 36% | 81% | 19% |
| 15 | The interviewer's ability to "ad lib" or deviate from a script during an introduction. | 2.95 (.89) | 29% | 72% | 28% |
| 16 | How high or low the interviewer's voice sounds (pitch). | 2.70 (.74) | 12% | 63% | 37% |
| 17 | The interviewer emphasizes the length of the survey. | 2.67 (.67) | 12% | 65% | 35% |
| 18 | The interviewer emphasizes the incentive. | 2.47 (.74) | 10% | 51% | 49% |

performance. Currently, most practitioners train and encourage interviewers to follow a script and, to a lesser degree, emphasize “ad libbing” during an introduction. It appears that practitioners recognize the importance of not sounding scripted; however, unscriptedness is admittedly difficult to train and measure.

Recommendations for interviewer training can be made based on these results. Specifically, interviewer response rates may benefit from an emphasis in interviewer training on speech that is as natural and unscripted as possible, through the use of intonation patterns and word selection. Interviewers can be exposed to contacts with both high and low ratings of scriptedness to make the difference clear.

While interviewers may be required to mention particular points in their introduction or even to follow a verbatim introductory script, they should be trained to sound as conversational as possible, particularly at the start of their introduction. Both Houtkoop-Steenstra and van den Bergh (2000) and Morton-Williams (1993) found that interviewers who were allowed to adapt their introductory script had greater success.

Beyond the introduction, the issue of standardized interviewing, and what departures from verbatim interview scripts can mean for data quality, is the subject of much debate. Schober and Conrad (1997; Conrad and Schober 2000) found clear evidence that “conversational” interviewing, or allowing interviewers to use any means necessary to convey question meaning, can enhance data accuracy. Nevertheless, “reading the questions exactly as worded” is a tenet of interview administration which is upheld and enforced in most survey organizations, and it is clear from results of the practitioners’ study that standardized interviewing skills are a high priority in nearly all organizations. Because emphasizing the need to read questions in a standardized manner may seem to conflict with emphasis on less-scripted delivery of introductions, interviewers need to be trained to “wear two hats.” In training, it needs to be made explicit to interviewers that there are two distinct (but, arguably, equally important) elements of the phone component of their job, each requiring a different style of speech and interaction. In the introductory or persuasive portion, scriptedness may be a liability, and the ability to “think on one’s feet” to respond to answers is an asset. In contrast, in the interviewing portion, deviating from a script may have ramifications for data quality, or at the very least will represent a lack of adherence to the organization’s procedures. Interviewers should be trained to “switch gears” between these two speech styles, and perhaps even be encouraged to acknowledge to respondents that their delivery of the questions will sound different from their introduction.

Finally, additional research could further these findings. Interviewer-level analyses, such as a larger study collecting ratings of characteristics for a greater number of contacts per interviewer to measure the impact of ratings on overall success rates, is recommended. Replicating the listeners’ study using a greater number of contacts by the same interviewers may shed light on those interviewer characteristics or behaviors across multiple contacts that lead to greater success.

5. References

Asch, S.E. 1946. “Forming Impressions of Personality.” *Journal of Abnormal and Social Psychology* 9: 258–290. Doi: <http://dx.doi.org/10.1037/h0055756>.

- Benkí, J., J. Broome, F. Conrad, R. Groves, and F. Kreuter. 2011. "Effects of Speech Rate, Pitch, and Pausing on Survey Participation Decisions." Paper presented at the 66th annual conference of the American Association for Public Opinion Research, Phoenix, AZ, May 14. Available at: <http://www.amstat.org/sections/srms/proceedings/42011/files/400189.pdf> (accessed September 29, 2015).
- Cialdini, R.B. 1984. *Influence: Science and Practice*. New York: Harper Collins.
- Conrad, F., J. Broome, J. Benkí, R. Groves, F. Kreuter, D. Vannette, and C. McClain. 2013. "Interviewer Speech and the Success of Survey Invitations." *Journal of the Royal Statistical Society* 176: 191–210. Doi: <http://dx.doi.org/10.1111/j.1467-985X.2012.01064.x>.
- Conrad, F.G. and M.F. Schober. 2000. "Clarifying Question Meaning in a Household Telephone Survey." *Public Opinion Quarterly* 64: 1–28. Doi: <http://dx.doi.org/10.1086/316757>.
- Curtin, R., S. Presser, and E. Singer. 2005. "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69: 87–98.
- Fiske, S.T., A.J.C. Cuddy, and P. Glick. 2007. "Universal Dimensions of Social Cognition: Warmth and Competence." *Trends in Cognitive Science* 11: 77–83. Doi: <http://dx.doi.org/10.1016/j.tics.2006.11.005>.
- Groves, R.M., B.C. O'Hare, D. Gould-Smith, J. Benkí, and P. Maher. 2007. "Telephone Interviewer Voice Characteristics and the Survey Participation Decision." In *Advances in Telephone Survey Methodology*, edited by J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. de Leeuw, L. Japac, P.J. Lavrakas, M.W. Link, and R.L. Sangster. (pp. 385–400). New York: Wiley.
- Groves, R.M., S. Presser, and S. Dipko. 2004. "The Role of Topic Interest in Survey Participation Decisions." *Public Opinion Quarterly* 68: 2–31. Doi: <http://dx.doi.org/10.1093/poq/nfh002>.
- Houtkoop-Steenstra, H., and H. van den Bergh. 2000. "Effects of Introductions in Large-Scale Telephone Survey Interviews." *Sociological Methods and Research* 28: 281–300. Doi: <http://dx.doi.org/10.1177/0049124100028003002>.
- Huefken, V. and A. Schaefer. 2003. "Zum Einfluss Stimmlicher Merkmale und Ueberzeugungsstrategien der Interviewer auf die Teilnahme in Telefonumfragen." *Koelner Zeitschrift fuer soziologie und sozial psychologie* 55: 321–339.
- Kelley, H.H. 1950. "The Warm-Cold Variable in First Impressions of Persons." *Journal of Personality* 18: 431–439. Doi: <http://dx.doi.org/10.1111/j.1467-6494.1950.tb01260.x>.
- Ketrow, S.M. 1990. "Attributes of a Telemarketer's Voice and Persuasiveness: A Review and Synthesis of the Literature." *Journal of Direct Marketing* 4: 7–21. Doi: <http://dx.doi.org/10.1002/dir.4000040304>.
- Morton-Williams, J. 1993. *Interviewer Approaches*. Cambridge: Cambridge University Press.
- Oksenberg, L. and C. Cannell. 1988. "Effects of Interviewer Vocal Characteristics on Nonresponse." In *Telephone Survey Methodology*, edited by R.M. Groves, P.B. Biemer, L.E. Lyberg, J.T. Massey, W.L. II, Nichols, and J. Waksberg. (pp. 257–272). New York: John Wiley and Sons.

- Oksenberg, L., L. Coleman, and C. Cannell. 1986. "Interviewers' Voices and Refusal Rates in Telephone Surveys." *Public Opinion Quarterly* 50: 97–111. Doi: <http://dx.doi.org/10.1086/268962>.
- O'Muircheartaigh, C. and P. Campanelli. 1999. "A Multilevel Exploration of the Role of Interviewers in Survey Nonresponse." *Journal of the Royal Statistical Society Series A* 162: 437–446. Doi: <http://dx.doi.org/10.1111/1467-985X.00147>.
- Rosenberg, S., C. Nelson, and P.S. Vivekananthan. 1968. "A Multidimensional Approach to the Structure of Personality Impressions." *Journal of Personality and Social Psychology* 9: 283–294. Doi: <http://dx.doi.org/10.1037/h0026086>.
- Schober, M.F. and F.G. Conrad. 1997. "Does Conversational Interviewing Reduce Survey Measurement Error?" *Public Opinion Quarterly* 61: 576–602.
- Sharf, D.J., and M.E. Lehman. 1984. "Relationship Between the Speech Characteristics and Effectiveness of Telephone Interviewers." *Journal of Phonetics* 12: 219–228.
- Snijders, G., J. Hox, and E.D. de Leeuw. 1999. "Interviewers' Tactics for Fighting Survey Nonresponse." *Journal of Official Statistics* 15: 185–198.
- Steinkopf, L., G. Bauer, and H. Best. 2010. "Nonresponse in CATI-Surveys." *Methods, Data, and Analysis* 4: 3–26. Available at: http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.4_Heft_1/01_Best.pdf (accessed September 29, 2015).
- Teitler, J.O., N.E. Reichman, and S. Sprachman. 2003. "Costs and Benefits of Improving Response Rates for a Hard-to-Reach Population." *Public Opinion Quarterly* 67: 126–138.
- United States Census Bureau. 2010. "Educational Attainment." Available at: <http://www.census.gov/hhes/socdemo/education/data/cps/2010/tables.html>.
- Van der Vaart, W., Y. Ongena, A. Hoogendoorn, and W. Dijkstra. 2005. "Do Interviewers' Voice Characteristics Influence Cooperation Rates in Telephone Surveys?" *International Journal of Opinion Research* 18: 488–499. Doi: <http://dx.doi.org/10.1093/ijpor/edh117>.

Received November 2012

Revised February 2015

Accepted March 2015