

# The Effects of Digital Recording of Telephone Interviews on Survey Data Quality

Danna Basson  
University of Wisconsin Survey Center

**Key Words:** Digital Recording, Participation, Satisficing, Data Quality

## 1. Introduction

Because digital recording is relatively unobtrusive, we assume it does not affect interviewer-respondent interaction and that recordings capture exactly how the interview would have proceeded otherwise. However interviewers may behave differently knowing they are recorded. Also, the technology is not completely unobtrusive and interviewers notice differences in the flow of the interview as recordings are saved on the computer. Finally, we also have to ask respondents for permission to record, which may affect their willingness to participate or their responses.

The aim of this research is to examine how recording affects survey data quality. Sample cases in a public opinion study were randomly assigned to be recorded or not in order to isolate potential effects of the recording on participation rates and data quality. The primary outcome measures related to survey participation examined here include completion and refusal rates, break-offs during the interview, the time it takes an interviewer to convince a respondent to participate, and the time to completion.

Since recording might also affect the type of responses people give, this paper also examines whether there are differences in survey satisficing (Krosnick, 1991) depending on whether the respondent is being recorded or not. The outcomes examined here include whether respondents who are being recorded provide more thoughtful answers. Measures of thoughtfulness include indicators of acquiescence bias, social desirability bias, responses to open-ended items, correct answers to factual questions, non-differentiation of answers, and item nonresponse.

## 2. Background

Recording of interviews can provide data that would otherwise be difficult or impossible to obtain. For example, researchers are using recordings to gather cognitive data previously only gathered in psychology labs. The current wave of the *Wisconsin Longitudinal Study* and the national *Midlife in the US Study* both conduct standard cognitive tests over the phone, such as category and letter fluency tasks, digit ordering tasks, and immediate and delayed recall tests, among others.

For some of these tasks, recording provides a check on the interviewer's data entry. For others such as letter fluency—where the respondent has one minute to name as many words as possible that begin with a given letter—the interviewer cannot type the entire response accurately. Instead, coders listen to recordings after the interview and code the respondent's answers.

Recording is also used to collect response times. Depending on the CATI software used, response latency can be computed using latent timers that log the total amount of time the interviewer and respondent spend at each question. Response latency can be timed more directly by having interviewers activate a timer with a keystroke immediately after reading a question, and end the timer with a second keystroke as soon as the respondent answers. Bassili and Fletcher (1991) also collect latency data using a voice-key, an electronic device that is connected to the phone and picks up the respondent's first utterance.

These methods provide useful data but can be extremely error prone. The interviewer-respondent interaction can result in latencies that are "contaminated" by respondent questions, hesitations, or "trigger happy" interviewers (Bassili 2000). This can result in a loss of latency data for between 11% of the cases (Bassili and Scott 1996) to over 30% of cases (Bassili and Fletcher 1991). Recording and coding post hoc provides an opportunity to collect this information more reliably.

Moreover, recording provides data for researchers who study interviewer-respondent interaction. The availability of software for analysis of such interactions facilitates the expansion of research in this area (Dijkstra 1999). For example, it can provide a rigorous method of evaluating survey questions during a pretest instead of relying on interviewer debriefings, which provide primarily anecdotal information (Oksenberg, Cannell and Kalton 1991). Behavior coding can also reveal how interviewers persuade reluctant respondents to participate in surveys (Dijkstra and Smit 2002). Such data has also been used to examine how interviewer behavior such as probing, providing feedback, maintaining objectivity, developing rapport, or other departures from the rules of standardized interviewing affect respondent behavior and the quality of responses (e.g. van der Zouwen, Dijkstra, and Smit, 1991).

Thus, recording can be used to improve the quality of surveys in a variety of ways. While interviews were traditionally recorded with cassette tapes that had to be

manually turned on by interviewers, recent advancements in digital technology make recording considerably less intrusive. In a CATI interview, recording can start automatically and the digital data is saved to the computer without interviewer involvement. The ease of digital recording makes the issue of how it affects the phone interview potentially relevant to more researchers.

There are good reasons to expect that recording might affect data quality. Recording might affect respondent behavior in ways that affect their willingness to participate as well as the quality of their answers, either in terms of the substantive content or the thoughtfulness of their answers. Recording might also affect interviewer behavior in conscious or unconscious ways, such as reading questions more slowly, repeating answer categories in full rather than just in part, and tensing up or preparing before the interview begins.

The exact way in which recording affects either respondents or interviewers cannot be isolated with this current dataset since the behavior of one party probably affects the behavior of the other as well. For example, the pace at which the interviewer reads the question might affect the thoughtfulness of the respondent's answer. Thus I can only observe the end result of the interaction of any of the above factors.

## 2.1 Survey Participation

Groves and Couper (1998) note that when respondents are approached with a request for participation in a survey, they quickly try to understand the nature of the request from the interviewer. Interviewer training often addresses (either explicitly or implicitly) the need by interviewers to discern and quickly address the nature of the respondent's concerns about participation. They do this by trying to make more or less salient the positive or negative reasons for participating in the survey (Groves, et al 2000). It is possible that recording might activate respondents' negative predispositions towards survey participation by calling attention to confidentiality concerns. If interviewers feel more nervous when they are being recorded, they might not be as effective on the phone.

It is also possible that recording might motivate respondents to participate or interviewers to avert refusals as it calls attention to the importance of the data to the researchers. If this is true, we would see a higher cooperation rate when recording. However the expectation here is that in most RDD surveys, asking for permission to record inhibits the interviewer and motivates the respondent to refuse rather than participate in the survey. Though one might expect that the ease of digital recording would make the

process inconspicuous, debriefings with interviewers after the field period for this project suggested that they found recording to be intrusive. Many complained of its effect on the interview (for example, they noted pauses and long delays) and were certain that refusals among respondents were considerably higher as a result of the recording. Thus anecdotal evidence at least suggests that recording inhibits participation.

It would be important to know if recording affects either the number or type of respondents who participate in a survey. These affect both the cost of obtaining the interviews as well as the representativeness, and ultimately the generalizability, of the data obtained.

## 2.2 Data Quality

If more willing survey participants tend to provide more thoughtful answers (Holbrook et al 1998, 2003) and one result of recording is that the ultimate sample of respondents is composed of people who are more willing to participate, this might result in observable differences in the data between interviews that are recorded and not.

Recording might also affect the quality of respondents' answers more directly. First, it is possible that if respondents feel like someone will be listening to their interview, they might be more careful in their answers. Secondly, there is some evidence to suggest that when the importance of their answers is underscored through various means, respondents will give more accurate data. Recording might imply to respondents that their responses are important, thereby motivating them to provide better answers.

It is also possible that asking for permission to record highlights the presence of the interviewer or the researcher. If this is true, then differences in data quality between interviews that are recorded and not might resemble the differences in data quality between telephone and in-person interviews, if to a smaller degree. For example, there is some evidence to suggest that in-person interviews about politics result in higher quality (or more thoughtful) responses (Green, Krosnick and Holbrook, 2001). These differences might also be observed when recording.

In this paper, I use the framework provided by Krosnick's (1991) theory of survey satisficing to evaluate the quality of data provided by respondents. The model of satisficing is based on the assumption that providing the optimal answer to a survey question involves a lot of cognitive work. The survey response process requires that a respondent interpret the survey question, engage in a memory search to retrieve information, integrate that information into a summary judgment, and then map that judgment onto the

response categories provided in the question (Tourangeau 1984).

According to Krosnick's theory, respondents might engage in satisficing, or reduce the amount of cognitive effort they expend answering survey questions. The likelihood and degree of satisficing depends on respondent ability, respondent motivation, and the difficulty of the task involved. It is possible that recording affects respondent motivation, making respondents in the recording condition more likely to give more thoughtful answers. This might be due to respondents feeling like their answers are more likely to be noticed, or because interviewers are more careful when they are being recorded, thereby motivating respondents in the process. For example, if recording motivates interviewers to slow down their pace, it might encourage respondents to give more careful answers, resulting in an increase in data quality when recording.

There are several observable implications of lower data quality. This might include increased acquiescence bias, or the tendency to agree with an assertion regardless of its content. It might also increase no-opinion responding, which is observable by the proportion of 'don't know' and 'refused' responses. Another discernible implication is willingness to give answers to open-ended questions, especially if those questions come soon after the request for permission to record. Also, if respondents are listening and responding more carefully when they are being recorded, then there should be a difference in the distribution of responses to knowledge or factual questions in the survey. Finally, it might also include nondifferentiation in responses. This occurs when a respondent chooses the same point on a scale for every consecutive question that uses that scale. Rather than retrieve information from memory, a respondent might choose a response that is easy to defend, possibly in the middle of a scale, regardless of the content of the question.

Finally, it is also possible that recording affects data quality in the opposite direction. That is, for certain types of questions, it is possible that recording encourages respondents to provide less accurate data by encouraging them to provide a socially desirable response. Thus it is also useful to compare responses to questions about socially desirable behaviors.

### 3. Data and Methods

The data for this analysis come from the *Badger Poll*, a statewide random digit dial survey conducted by the University of Wisconsin Survey Center in June of 2004. The field period for interviewing 500 respondents for this survey is generally one week to

ten days; the data for this analysis were collected between June 15 to 23, 2004.

The *Badger Poll* is primarily a public opinion survey with questions about national and Wisconsin politics and culture. The content of this Poll included evaluations of President George W. Bush, Wisconsin Governor Jim Doyle, Wisconsin Senators Herb Kohl and Russ Feingold; issue questions about state budget, state tax policy, the war in Iraq and the Iraqi prisoner abuse scandal; and evaluations of the candidates running for President and for a WI senatorial seat. The average time for a complete interview was 26.9 minutes.

Occasional administrations of the *Badger Poll* also contain tack-on questions by external clients. The Poll data used for this analysis contains questions that were asked at the end of the survey by the Wisconsin Stroke Alert. This series contains twenty two questions primarily about knowledge of stroke symptoms. These questions accounted for an average of 2.51 minutes of the total time.

The request for permission to record the interview occurs immediately after the interviewer reads the confidentiality statement to the respondent. Respondents are told that their participation is voluntary and that their information and responses are confidential and will only be used for statistical purposes. The permission to record is asked as follows: "For research purposes, I would like to record this interview. Is that OK with you?" Interviewers were instructed to treat this question like any other on the survey and accept the respondent's answer without trying to persuade them otherwise. Of the 504 completed interviews, 354 (or 70%) of them were in the recorded condition. Of those 354, only 9 respondents (or 2.5%) refused to be recorded.

## 4. Results

### 4.1 Survey Participation

Because recording might affect a respondent's likelihood of participating indirectly via the interviewer's behavior, it is necessary to examine cooperation at different key points at the start of a phone call. The points available in this dataset include: prior to determining if the number dialed belongs to an eligible household; after requesting permission to record; and after the substantive questions in the interview have begun.

There are no statistically significant differences in the number of completes, refusals, or other terminations based on recording (see Table 1). If there were differences in interviewer behavior that may have caused respondents to be more or less likely to participate in the survey prior to being asked for

permission to record, they do not result in differential participation rates in this data. Differences in rates between the recorded and not recorded condition are fewer than 3% points, well under the standard error of the mean for every measure reported here. There are no cases where the selected respondent refused as a direct result of the request to record. Furthermore, there is no discernable tendency in the direction of the results.

In addition to the rates at which people ultimately choose to participate in the survey, calling effort might be affected by recording. If recording activates a person's general unwillingness to cooperate with the survey request, it is possible that it might require additional phone calls to the household to convince the respondent to participate. The time it takes to introduce the study and read through the confidentiality statement might also take longer if interviewers have to spend more time explaining the study to respondents, emphasizing the legitimacy of the survey organization, or justifying the purpose of recording. If recording increases reluctance to participate in the survey, the costs associated with administering the survey might increase even if the ultimate participation rates are the same in both conditions. However, I find no statistically significant differences in willingness to participate on these measures when looking at introduction time and at the number of call attempts per interview by recording condition (see Table 2).

Besides cooperation rates, it is possible that recording interacts with other respondent characteristics, such as sex, age, and race. These demographic characteristics are related to numerous aspects of survey participation (Groves and Couper 1998). For example, we know that women, older respondents, higher-income respondents, and white respondents are generally more likely to cooperate with a survey request. Since this dataset comes from a RDD survey, we cannot know this information for nonrespondents, however if any of these characteristics interact with recording, then we would see differences in the distribution of these characteristics by recording condition. Consistent with the results shown above, there are no differences in the demographic makeup of respondents in both conditions (see Table 3).

## 4.2 Data Quality

Besides the potential effects of recording on data quality as a result of its potential effect on participation and representativeness, it is possible that recording affects the quality of the data that willing respondents give over the phone when they know they are being recorded. As shown above, there are no differences in demographic characteristics of respondents in each

recording condition. Thus if there are any differences in data quality in each condition, this should be due to the effect of recording and not to any effect recording might have on survey participation. The key outcomes used to examine data quality are some of the observable implications of survey satisficing, namely: acquiescence bias, item nonresponse, correct answers to knowledge or factual questions, whether respondents provided responses to open-ended items, and nondifferentiation in responses.

If respondents are satisficing, they should be more likely to acquiesce to questions in the affirmative. This is measured by counting the number of 'yes' responses to yes/no questions on the Poll. There were twenty such questions on the Poll. This indicator measures the proportion of these questions that were answered with a response of 'yes' and coded to range between 0 and 1. (All of the indicators reported here were coded to range between 0 and 1 unless otherwise noted). There were no significant differences in acquiescence by recording condition (see Table 4 in Appendix A).

If recording makes respondents more self-conscious, they should be more likely to provide a socially desirable response under the recording condition. There were six questions where social desirability might be a factor, including whether the respondent pays attention to politics, is registered to vote, and is aware of stroke symptoms. After counting the proportion of responses where the respondent provided the socially desirable answer, there were no significant differences in social desirability by recording condition (Table 4). Though such questions are commonly thought to be susceptible to social desirability effects (Holbrook et al 2003), they might not have had an effect among this population.

If respondents are more thoughtful as a result of recording and searching their memory to provide a more accurate answer, then they should also be more likely to provide a correct answer to factual questions. There were three questions about political knowledge that asked who the Vice President of the US is, which party is more conservative, and which party controls the WI legislature – all standard items used to measure political knowledge. There were no differences in the proportion of correct answers by recording condition. In addition, there were twenty items asking respondents if certain conditions were symptoms of a stroke (such as dizziness, numbness in the face, chest pain, etc). There were no differences in the proportion of correct items here either (Table 4).

If recording makes respondents more thoughtful, they should be less likely to provide a no-opinion response such as a 'Don't Know' or a 'Refusal' when they are being recorded. Looking at the mean number of no-opinion responses, there are no significant

differences by recording condition. The tendency is in the expected direction; that is, recording results in approximately one fewer no-opinion response (Table 4). There were also two open-ended items on the Badger Poll, asking respondents what they think is the most important problem in the US today, and the most important problem in WI today. These are the second and fifth question in the survey, respectively. Since these two questions appear relatively soon after requesting permission to record, they might be more likely to show evidence of a recording effect. However there are no significant differences in the proportion of responses to either question.

Finally, looking at nondifferentiation, I find consistent results with the patterns above, namely no effect of recording (Table 4). I looked at three different series of questions and counted the number of instances where the respondent provided the same response on the scale for consecutive items. Recording has no effect on this measure.

These measures of satisficing might not be evident among respondents in general but might be more evident among groups that are more vulnerable to satisficing (or strong satisficing). Thus I examined all of the above measures by various demographic characteristics, namely education level, age, sex, race, household income, and whether the respondent pays attention to politics since this is a political survey. While some of these variables explain differences in satisficing, recording has no effect at all after controlling for these variables. Finally, in case the differences were too small to detect without holding all of the other variables constant, I ran multivariate regression on all of these measures and recording still has no effect on any of the satisficing indicators.

## 5. Conclusion

Despite expectations to the contrary, recording does not appear to have an effect on willingness of potential respondents to participate in a statewide RDD CATI interview about political affairs. Potential effects on cost or data quality by way of representativeness do not appear to be at issue. Furthermore, effects on the quality of data provided by respondents also do not appear to be an issue.

It is possible that the evidence showing no effect of recording on data quality might be related to the substantive topic of the survey. The *Badger Poll* is primarily a survey of attitudes about politics. Given the relatively low levels of political interest and knowledge in the general population, the survey is of relatively low salience to many people. A topic of greater concern to the sample population might show different results.

Furthermore, none of the survey topics were potentially sensitive issues. Recording is more likely to affect both willingness to participate and interviewer-respondent interactions in a survey about sensitive behaviors, where concerns about confidentiality or interviewer feedback might be more salient. In such a situation, recording could interact with characteristics of interest. However this pattern is not evident in this data.

That there are no significant differences in participation rates between the recorded and not-recorded conditions is contrary to expectations. During interviewer debriefings for the *Badger Poll*, many interviewers reported being very aware of the recording. Their anecdotal impressions from having called this project were that recording made it difficult to keep respondents on the phone and they were fairly certain that recording contributed to a higher number of refusals. This is not borne out in the data. Instead, recording appears to have no effect on the rates at which people were willing to participate in the survey.

Since recording can be used to improve the quality of surveys by monitoring interviewer performance and analyzing pretest interviews, the ease of digital recording makes the issue of how it affects the phone interview potentially relevant to more researchers. That recording has no effect is a positive outcome. Since digital technology makes it a relatively simple matter to record interviews, this might become a more commonly used tool by researchers. It might be necessary to address interviewers' concerns prior to recording, since it is primarily interviewer response to the recording that inspired this particular paper. However the evidence presented here suggests that despite interviewer perceptions, recording interviews does not have adverse results on survey participation or data quality on a political survey.

## References

- Bassili, John N and Fletcher, Joseph F, 1991. "Response-Time Measurement In Survey Research." *Public Opinion Quarterly*, 55(3):331-347.
- Bassili, J N and Stacey Scott, 1996. "Response Latency as a Signal to Question Problems in Survey Research." *Public Opinion Quarterly*. 60(3): 390-399.
- Dijkstra, Wil, 1999. "A New Method for Studying Verbal Interactions in Survey Interviews." *Journal of Official Statistics*, 15(1): 67-85.
- Dijkstra, Wil and Johannes H Smit (2002). "Persuading Reluctant Recipients In Telephone Surveys." In *Survey Nonresponse*. Robert M. Groves,

Don A Dillman, John L. Eltinge, Roderick JA Little (eds). Wiley. New York.

Ellis, Charles H. and Jon A. Krosnick. 1999. "Comparing Telephone and Face-to-Face Surveys in Terms of Sample Representativeness: A Meta-Analysis of Demographic Characteristics." NES Technical Report Series, No. nes010871

Groves, Robert m, F Fowler, M Couper, J Lepkowski, E Singer, and R Tourangeau, 2004. *Survey Methodology*. John Wiley and Sons, Inc. New Jersey.

Groves, Robert and Mick Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: Wiley.

Groves, Robert; Singer, Eleanor; Corning, Amy, 2000. "Leverage-Saliency Theory of Survey Participation." *Public Opinion Quarterly*. 64(3):299-309

Green, Melanie C, Jon A Krosnick, and Allyson L. Holbrook, 2001. "The Survey Response Process in Telephone and Face-to-Face Surveys: Differences in Respondent Satisficing and Social Desirability Response Bias." <http://communication.stanford.edu/faculty/krosnick>

Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone Versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondents

Satisficing and Social Desirability Response Bias." *Public Opinion Quarterly*. 67:79-125.

Krosnick, J A, 1991. "Response strategies for coping with the cognitive demands of attitude measures in surveys." *Applied Cognitive Psychology*. 5:213-236

Oksenberg, L, C Cannell, and Graham Kalton, 1991. "New Strategies for Pretesting Survey Questions." *Journal of Official Statistics*. 7(3):349-365.

Tourangeau, Roger, 1996. "Cognitive Sciences and Survey Methods." In *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Eds Thomas B Jabine, Miron Straf, Judith Tanur and Roger Tourangeau. Washington, DC: National Academy Press.

van der Zouwen, Johannes, Wil Dijkstra, and Johannes H Smit, 1991. "Studying Respondent-Interviewer Interaction: The Relationship Between Interviewing Style, Interviewer Behavior, and Response Behavior." Pp 419-437 in *Measurement Errors in Surveys*, edited by P. P. Biemer, R. M. Groves, L.E. Lyberg, N. A. Mathiowetz and S. Sudman. New York: John Wiley & Sons, Inc.

Wisconsin Longitudinal Study (WLS) [graduates]: 2003/5. Hauser, Robert M.; Hauser, Taissa S. Principal investigators. Madison, WI: University of Wisconsin-Madison, WLS. <http://www.ssc.wisc.edu/~wls>

**TABLES**

**Table 1: Cooperation Rates, by Recording Condition**

	Recording Condition		Difference
	Recording	Not Recording	
<b>Proportion of Numbers of Unknown Eligibility</b>	4.8% N=67	3.6% N=23	+ 1.2%
<b>Breakoffs After Start of Interview (% among started interviews)</b>	9.9% N=39	10.2% N=17	- 0.3%
<b>Completes</b>	23.36% N=355	25.61% N=150	- 2.25%
<b>Refusals</b>	32.55% N=474	34.20% N=209	- 1.64%
<b>Completes/ (Completes + Refusals)</b>	0.428	0.418	+ 0.01

Differences not statistically significant

**Table 2: Mean Introduction Time, Interview Time, and Call Attempts, by Recording Condition**

	Recording Condition		Difference
	Recording	Not Recording	
<b>Mean Introduction Time</b>	88.67 secs N=340	90.64 secs N=164	- 1.97 seconds
<b>Mean Total Interview Time</b>	26.76 mins N=355	27.39 mins N=150	- 0.63 minutes
<b>Mean Number of Call Attempts</b>	5.61 calls N=354	6.09 calls N=150	- 0.48 calls

Differences not statistically significant

**Table 3: Demographic Characteristics by Recording Condition**

	Recording Condition		Difference
	Recording	Not Recording	
<b>Percent Female</b>	52.3%	55.3%	- 3.0%
<b>Mean Respondent Age</b>	48.5 years	50.3 years	- 1.8 years
<b>Proportion of Non-white Respondents</b>	8.5%	6.0%	+ 2.5%

Differences not statistically significant

**Table 4: Satisficing Indicators, by Recording Condition**

	Recording Condition		Difference
	Recording	Not Recording	
<b>Acquiescence</b>			
Mean proportion of 'yes' responses to yes/no questions	0.614	0.624	- 0.010
<b>Social Desirability</b>			
Proportion giving the socially desirable	0.292	0.302	- 0.010
<b>Knowledge</b>			
Mean proportion of correct answers to questions about politics	0.593	0.627	- 0.034
Mean proportion of correct answers to questions about stroke symptoms	0.675	0.661	+ 0.014
<b>No Opinion</b>			
Mean Number of Items with a DK or REF Response	8.47	9.56	- 1.09
<b>Mean proportion providing both open-ended responses</b>	0.935	0.937	- 0.002
Most important problem, US	0.969	0.980	- 0.011
Most important problem, WI	0.898	0.893	+ 0.005
<b>Nondifferentiation: Mean proportion of like responses</b>			
9-question series evaluating Bush and			

AAPOR - ASA Section on Survey Research Methods

Kerry on several attributes	0.698	0.694	+ 0.004
8-question series rating President, Senators, and Congress on policies	0.603	0.588	+ 0.015
5-question series evaluating the potential outcomes of a WI tax reform bill	0.597	0.577	+ 0.020

---

Data Source: *Badger Poll #16*  
Differences not statistically significant