

Item-Nonresponse and the 10-Point Response Scale in Telephone Surveys

Matthew Courser*, Paul J. Lavrakas†

Keywords: measurement error, survey error, scale construction

DOI: [10.29115/SP-2012-0021](https://doi.org/10.29115/SP-2012-0021)

Survey Practice

Vol. 5, Issue 4, 2012

Item-Nonresponse and the 10-Point Response Scale in Telephone Surveys

So-called “10-point” rating scales are one of most commonly used measurement tools in survey research and have been used successfully with many types of constructs including items that ask respondents to rate their satisfaction with political leaders, the economy, and with their overall quality of life. However, the exact format of the 10-point response scales used has varied widely with some researchers using scales that run from 1-10 and others using scales that run from 0-10. In addition, the number of scale points assigned labels varies with some researchers labeling only the endpoints, others labeling the endpoints and scale midpoint, and still others labeling all of the scale points.

Previous research (Andrews 1984; Cox 1980; Garratt et al. 2011; Schwarz et al. 1991) has sought to understand how response scales can influence the distribution of survey data and how the labeling and design of response scales influence the validity and reliability of survey data. Although the literature on response scales and their effects on survey data is extensive, scholars have yet to report investigations of the linkages between response scales and resulting item nonresponse. In particular, little is known about the impact of the format of the 10-point response scale on levels of item-nonresponse in survey data.

We seek to increase knowledge on this issue by reporting the results of two experimental studies that were designed to test whether the format of 10-point response scale used has a significant and nonignorable influence on item nonresponse and thus, on levels of data quality in RDD surveys. In doing so, we argue that when designing a 10-point scale, researchers must consider not only the validity and reliability of the scale, but also the level of anticipated item-nonresponse from the scale format.

* **Institution:** Pacific Institute for Research and Evaluation—Louisville Center

† **Institution:** Independent Consultant

METHODOLOGY

Two studies were conducted to understand the impact of 10-point scale format on survey data, with the first focusing on the relationship between scale format and candidate favorability ratings and the second focusing on scale format and economic ratings. The data come from two Buckeye State Polls conducted in 2000 and 2001 by the former Center for Survey Research at Ohio State University. The Buckeye State Poll was a statewide RDD survey of Ohio residents conducted monthly from 1996 to 2001.

Study 1 used Buckeye State Poll data ($n = 1,525$) that were collected from October 1, 2000, to October 31, 2000, to explore the impact of 10-point scale format on candidate favorability ratings for four candidates – Presidential candidates Al Gore and George W. Bush, and the United States Senate candidates in Ohio, Ted Celeste and Mike DeWine. The response rate (AAPOR RR1) for the October 2000 Buckeye State Poll data was 48%, and the cooperation rate (AAPOR COOP1) was 80%.

Study 2 used Buckeye State Poll data ($n = 1,153$) that was collected during March and April 2001 to explore the impact of 10-point scale format on approval or disapproval of three high-profile economic issues – (1) former President Bush’s proposed income tax cut, (2) Alan Greenspan and the Federal Reserve’s recent decisions to lower interest rates, and (3) former President Clinton’s plans to use the federal budget surpluses to reduce the national debt. The response rate (AAPOR RR1) for the March and April 2001 Buckeye State Poll data was 37%, and the cooperation rate (AAPOR COOP1) was 88%.

In both studies, respondents were randomly assigned to one of three conditions that manipulated the format of the scale used to make their ratings. For study 1 (candidate favorability), one condition used a 1–10 scale with 1 anchored with “very unfavorable,” and 10 with “very favorable.” Another condition was a 0–10 scale with 0 anchored with “very unfavorable” and 10 with “very favorable.” The third condition was a 0–10 scale with 0 anchored by “very unfavorable,” 5 anchored by “neither favorable nor unfavorable,” and 10 anchored by “very favorable.” For study 2 (economic issues), one condition used a 1–10 scale with 1 anchored with “strongly disapprove,” and 10 with “strongly approve.” Another condition was a 0–10 scale with 0 anchored with “strongly disapprove” and 10 with “strongly approve”. The third condition was a 0–10 scale with 0 anchored by “strongly disapprove”, 5 anchored by “neither approve nor disapprove,” and 10 anchored by “strongly approve”. In study 1, randomizations were used to vary the order of the candidates’ names and the order in which the two blocks of questions (i.e., the two presidential questions and the two Senate questions) were presented to respondents. In study 2, randomizations were used to vary the order of the economic issues.

RESULTS

Table 1 presents the average amount of nonresponse for the three versions of

the 10-point rating scales across both topic areas. Across the three conditions tested in each study, item nonresponse came from respondents saying “refused” or “don’t know” as their answers. For the two scales that included a true midpoint, Table 1 also reports on the proportion of the sample that used that midpoint (5) for at least one of the four candidate favorability or at least one of the three economic issue items.

Table 1 Average nonresponse across topic areas by scale type.

Response Scale Format	Average Nonresponse for 4 Candidate Favorability Items	Use of midpoint rating ("5") for 4 Candidate Favorability Items	Average Nonresponse for 3 Economic Issue Ratings	Use of midpoint rating ("5") for 3 Economic Issue Items
1-10 scale	19.4%	***	20.6%	***
0-10 scale	16.1%	17.3%	20.3%	52.6%
0-5-10 scale	11.7%	21.1%	13.9%	55.2%

*Midpoint rating not reported for 1-10 scale as "5" is not a true midpoint of the scale.

Table 1 shows that the groups of respondents who were used a 1–10 rating scale had the highest levels of item nonresponse (19.4% and 20.6%, respectively). The group of respondents that used a 0–10 rating scale had less item nonresponse (16.1% and 20.3%, respectively), and the group that used the 0–5–10 rating scale had the lowest levels of item nonresponse (11.7% and 13.9%, respectively). These differences in the proportion of item nonresponse across the three groups were statistically significant for both studies (study 1 chi-square: $df = 2$; $p < 0.05$; study 2 chi-square: $df = 2$; $p < 0.05$). Furthermore, the size of these differences is not ignorable, with the 1–10 scale having 66% more missing data than the 0–5–10 scale in study 1 and 48% more missing data in study 2.

In addition, Table 1 suggests that for the two scales that had a true midpoint (0–10 and 0–5–10), as item nonresponse decreased, there was a small trend toward respondents using the midpoint of the scale. These differences in the proportion of using the midpoint across the two groups receiving a scale with a true midpoint were statistically significant for both studies (study 1 chi-square: $df = 16$; $p < 0.05$; study 2 chi-square: $df = 9$; $p < 0.05$). This shift in the distribution of responses is not surprising, as people who are uncertain about an issue or candidate and who otherwise would contribute to item nonresponse by choosing “don’t know” would be expected to use a rating at the midpoint of the scale.

DISCUSSION

Using results from two sets of multi-item experiments, we have shown that common response scales used in a wide variety of surveys are significantly related to the amount of item non-response. Specifically, we found that 1–10 response scales produced much higher levels of missing data than 0–10 scales and that 0–10 scales with 5 anchored as a midpoint consistently had the smallest amount of data missing. These results were consistent across multiple surveys in both political and economic domains and across items that asked respondents to make favorability ratings and items that asked for approval judgments.

A key implication of our findings is that a 1–10 scale should not be used and that when designing a 10-point response scale for use with a telephone survey, a scale that runs from 0–10 and which has both the endpoints and the midpoint labeled will minimize item nonresponse.

A possible alternative explanation could be that our observed findings were not due to the impact of the scale configuration on respondents, but instead due to interviewer-related effects; 94 different interviewers worked on at least one of the studies, 22 of whom worked on both. That is, it is possible that some interviewers preferred one scale over the other and as such implemented the items differentially depending on which scale was being used. However, a series of analyses on whether there was any tendency within individual interviewers

to deviate from the general pattern of the findings showed no such support for that possible alternative explanation.

A larger question stemming from our results is why a 0–5–10 scale has lower levels of item nonresponse. We speculate that the 0–5–10 scale has lower levels of item nonresponse because it provides more information for respondents and thus helps them to make a “real” choice. Although we found some evidence that respondents used the midpoint value of “5” slightly more often when presented with a 0–5–10 scale, we believe that a “5” is a meaningful substantive response and that it is better to have respondents provide a valid (i.e., non-missing) value on the scale – from both a substantive and statistical standpoint – than to have to impute values for respondents with missing data.

Most important, our research extends previous work by showing that response scales influence item nonresponse – not just the answers or estimates provided by respondents.

NOTE

An earlier version of the results of this research study was presented at the 2001 Annual Meeting of the American Association for Public Opinion Research, May 17–20, Montreal, Quebec.

REFERENCES

- Andrews, F.M. 1984. "Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach." *Public Opinion Quarterly* 48: 409–42.
- Bendig, A.W. 1954. "Reliability and the Number of Rating Scale Categories." *The Journal of Applied Psychology* 38: 38–40.
- Bevan, W., and L. Avant. 1968. "Response Latency, Response Uncertainty, Information Transmitted and the Number of Available Judgemental Categories." *Journal of Experimental Psychology* 76: 394–97.
- Cox, E.P. 1980. "The Optimal Number of Response Alternatives for a Scale: A Review." *Journal of Marketing Research* 17: 407–22.
- Garratt, A.M., J. Helgeland, and P. Gulbrandsen. 2011. "Five-Point Scales Outperform 10-Point Scales in a Randomized Comparison of Item Scaling for the Patient Experiences. Questionnaire." *Journal of Clinical Epidemiology* 64 (2): 200–207.
- Gaskell, G.D., C. O’Muircheartaigh, and D. Wright. 1994. "Survey Questions about the Frequency of Vaguely Defined Events: The Effects of Response alternative." *Public Opinion Quarterly* 58: 241–54.
- Jacoby, J., and M. Matell. 1971. "Three Point Likert Scales Are Good Enough." *Journal of Marketing Research* 8: 495–500.
- Komorita, S.S., and W.K. Graham. 1965. "Number of Scale Points and the Reliability of Scales." *Educational and Psychological Measurement* 4: 987–95.
- Krosnick, J.A., and L. Fabrigar. 1997. "Designing Rating Scales for Effective Measurement in Surveys." In *Survey Measurement and Process Quality*, edited by Lyberg et al. New York: John Wiley & Sons.
- Lissitz, R.W., and S.B. Green. 1975. "Effect of the Number of Scale Points on Reliability: A Monte Carlo Approach." *Journal of Applied Psychology* 60: 10–13.
- Matell, M.S., and J. Jacoby. 1971. "Is There an Optimal Number of Alternatives for Likert Scale Items? Study I: Validity and Reliability." *Educational and Psychological Measurement* 31: 657–74.
- Miller, G.A. 1956. "The Magical Number Seven, plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63: 81–97.
- Schwarz, N. 1985. "Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgements." *Public Opinion Quarterly* 49: 388–95.
- . 1991. "Rating Scales: Numeric Values May Change the Meaning of Scale Labels." *Public Opinion Quarterly* 55: 570–82.