Proposal for the ANES 2016 Pilot Study: Improved measurement of feeling thermometer questions

Keywords:

Feeling thermometer; methodological experiment; randomized experiment

Research questions and motivations

The feeling thermometer is a unique question type that is frequently used in general population surveys and political polls to measure respondents' feelings toward a certain object, such as a political party or a politician. The feeling thermometer questions have traditionally been included in the ANES time series studies. It uses a 101-point rating scale, where 0 indicates very cold and unfavorable feeling while 100 indicates very warm and favorable feeling. The belief that one's feelings can affect one's behaviors, like voting, has contributed to the popularity of this question type (Buell & Sigelman, 1985; Greene, 1999, 2004; Kaid, Leland, & Whitney, 1992; Lauderdale, 2010; McAdams & Johannes, 1988).

Given the popularity of the feeling thermometer, the measurement of this question has also received some research attention (Chang & Krosnick, 2009; Green, 1988; Malhotra & Krosnick, 2007; Smith, 1987; Wilcox, Sigelman, & Cook, 1989). For example, in one study the test-retest reliability for the 7-point scale has been found to be higher than that of the 101-point scale (Krosnick, 1991b). Another study compares feeling thermometer questions in face-to-face surveys and Web surveys using the 2014 ANES time series studies (Liu & Wang, 2015). It shows that face-to-face respondents tend to provide warmer and more favorable responses, less precise or more rounded

answers, and less reliable responses than Web respondents. More recently, we conducted a Web survey experiment through TESS (Time-Sharing Experiments for the Social Sciences) to test six variations of feeling thermometer questions, including four types of visual analog scales (VAS), one drop-down menu, and one open-end numeric text box. The precision of responses from VAS is higher than that of numeric text box, which is the format currently used in ANES.

Verbal and numeric labels for feeling thermometer questions is one design feature that has not received much attention. Figure 1 presents the feeling thermometer question illustration used for the 2012 ANES time series studies. As can be seen, nine verbal and numeric labels are provided in the feeling thermometer, all of which are on degrees that are multiples of 5. Liu and Wang (2015) show that for each of the 45 feeling thermometer questions asked in the 2012 ANES, over 90% of face-to-face respondents selected one of these nine options. Although the effect of labels is slightly lower for the Web respondents, for most questions still over 80% of the responses are for one of these nine options. This suggests that the rounded answers are largely driven by the use of verbal labels. If we take rounding as a form of satisficing (Krosnick, 1991a), it is a clear sign of poor measurement quality as the original hope for adopting a 101-point rating scale is to enhance the level of granularity for survey responses. The reality is that, rather than using the full range of the rating scale, respondents focus on the response options with labels. This is likely due to the fact that labeled responses impose a lower cognitive burden for interpretation than unlabeled ones where the respondents have to assign meanings to them. The goal of this study is to test alternative question forms that may potentially reduce the reliance on labels and improve the precision of the responses. Specifically, the

primary research question is to examine whether responses to feeling thermometer questions are influenced by verbal labels. Also, this study will test whether different formats of feeling thermometer questions will result in different levels of response precision.

Figure 1. 2012 ANES Time Series Feeling Thermometer.

Design of experiment

This study proposes a single factor factorial experiment. It includes three experimental conditions:

- Condition 1 – Original ANES feeling thermometer question: Questions in this condition replicate the 2012 ANES question format and will be used as the control group.

- Condition 2 – Endpoint labeled feeling thermometer question: In this condition, the feeling thermometer questions will only include end labels, i.e., 0° Very cold or unfavorable feeling, and 100° Very warm or favorable feeling. No other labels are presented.

- Condition 3 – Off-multiples-of-5 labeled feeling thermometer question: the original ANES labels are moved down one degree, that is from 85°, 70°, 60°, 40°, 30° and 15° to 84°, 69°, 59°, 39°, 29° and 14°, respectively.

- Condition 4 – Numeric labeled feeling thermometer question: Remove the original ANES verbal labels but keep the numeric labels.

Respondents will be randomly assigned to one of these four conditions.

Question contents and wordings

This study will use the questions adapted from the 2012 ANES survey. The proposed lead-in and rating items are as follows:

We would like to know your feelings toward some political leaders and groups of people using something we call the feeling thermometer. Ratings between 50° and 100° mean that you feel favorable and warm toward the person or the group of people. Ratings between 0° and 50° mean that you don't feel favorable toward the person or group of people and that you don't care too much for them. You would rate the person or group of people at the 50° mark if you don't feel particularly warm or cold toward them.

How would you rate Barak Obama?

How would you rate Hillary Clinton?

How would you rate Donald Trump?

How would you rate the conservatives?

How would you rate the liberals?

How would you rate Congress?

How would you rate Big Business?

Analysis plan

Within each condition, two variables are calculated: the proportion of rounded answers (i.e., answers that are multiples of 5) and the proportion of answers that fall onto the labeled options. Comparisons between Conditions 1, 2, and 4 will answer the question of whether the answers are influenced by the verbal labels. Chi-square test will be performed. Comparison between Conditions 1 and 3 will answer the question of whether the rounded answers are due to verbal labels or numeric integers. Pairwise t-test will be performed. Variance in each condition will also be compared in order to determine in which condition the respondents are most likely to use a wider range of the scale. F-test will be performed. The mean of the feeling thermometer scores will be compared across all conditions, using chi-square test. Item non-response will also be tested across all conditions through chi-square test.

In addition to the overall comparison, similar analyses will be conducted by respondent's education and age. Previous research has shown that the impact of question format is not universal. Rather, it interacts with respondents' cognitive capacity (Knauper, 1999; Knäuper, Belli, Hill, & Herzog, 1997; Knauper, Schwarz, Park, & Fritsch, 2007). Since education and age are the two frequently used proxy measures for cognitive capacity, this study proposes to also use them to examine whether the impact of feeling thermometer questions interacts with cognitive capacity.

The time to complete the questionnaire will also be analyzed across the experimental conditions. It will be used as a proxy of task difficulty. A longer time may indicate a higher level of response difficulty.

Additional testing

In addition to the ANES study, we will also seek to conduct similar experiments using non-probability samples, and expand the variety of items. The comparison of findings between ANES and the non-probability sample is of great interest to researchers and will have methodological implications.

References

Buell, E. H., & Sigelman, L. (1985). An Army That Meets Every Sunday-Popular Support For The Moral Majority In 1980. *Social Science Quarterly*, *66*(2), 426– 434.

Chang, L., & Krosnick, J. A. (2009). National Surveys Via Rdd Telephone Interviewing Versus the Internet Comparing Sample Representativeness and Response Quality. *Public Opinion Quarterly*, *73*(4), 641–678.

Green, D. P. (1988). On the Dimensionality of Public Sentiment toward Partisan and Ideological Groups. *American Journal of Political Science*, *32*(3), 758–780.

Greene, S. (1999). Understanding Party Identification: A Social Identity Approach. *Political Psychology*, *20*(2), 393–403.

Greene, S. (2004). Social Identity Theory and Party Identification. *Social Science Quarterly*, *85*(1), 136–153.

Kaid, L. L., Leland, C. M., & Whitney, S. (1992). The Impact of Televised Political Ads: Evoking Viewer Responses in the 1988 Presidential Campaign. *Southern Communication Journal*, *57*(4), 285–295.

Knauper, B. (1999). The impact of age and education on response order effects in attitude measurement. *Public Opinion Quarterly*, 347–370.

Knäuper, B., Belli, R. F., Hill, D. H., & Herzog, A. R. (1997). Question difficulty and respondents' cognitive ability: The effect on data quality. *Journal of Official Statistics*, *13*(2), 181–199.

Knauper, B., Schwarz, N., Park, D., & Fritsch, A. (2007). The perils of interpreting age differences in attitude reports: Question order effects decrease with age. *Journal of Official Statistics*, *23*(4), 515-528.

Krosnick, J. A. (1991a). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236.

Krosnick, J. A. (1991b). The stability of political preferences: Comparisons of symbolic and nonsymbolic attitudes. *American Journal of Political Science*, 547–576.

Lauderdale, B. E. (2010). Unpredictable Voters in Ideal Point Estimation. *Political Analysis*, *18*(2), 151–171.

Liu, M., & Wang, Y. (2015). Data collection mode effect on feeling thermometer questions: A comparison of face-to-face and Web surveys. *Computers in Human Behavior*, *48*(7), 212–218.

Malhotra, N., & Krosnick, J. A. (2007). The effect of survey mode and sampling on inferences about political attitudes and behavior: Comparing the 2000 and 2004

ANES to Internet surveys with nonprobability samples. *Political Analysis*, *15*(3), 286–323.

McAdams, J. C., & Johannes, J. R. (1988). Congressmen, Perquisites, and Elections. *The Journal of Politics*, *50*(2), 412–439.

Smith, T. W. (1987). That Which We Call Welfare by Any Other Name Would Smell Sweeter an Analysis of the Impact of Question Wording on Response Patterns. *The Public Opinion Quarterly*, *51*(1), 75–83.

Wilcox, C., Sigelman, L., & Cook, E. (1989). Some Like It Hot: Individual Differences in Responses to Group Feeling Thermometers. *The Public Opinion Quarterly*, *53*(2), 246–257.