The effects of rating scale format on the measurement of policy attitudes in web surveys

Alexander Jedinger

GESIS – Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, 50667 Cologne, Germany

ARTICLE INFO

ABSTRACT

Recent research suggests that policy attitudes should be assessed with surveys that use branched rating scales that separately measure the direction and intensity of issue preferences. However, past studies have not tested whether branched rating scales improve the quality of survey data in the context of self-administered web surveys. In the current study, I compare the effects of different issue scale formats on response distributions, survey satisficing and issue-based voting in multiparty elections. Using data from a randomized web experiment, I find that partially labeled seven- and eleven-point issue scales and fully verbalized branched scales produce very similar response distributions and comparable effects of policy attitudes on party choice. However, the findings regarding the extent of satisficing behavior are mixed. Based on the efficiency of different scale formats, scholars are encouraged to use seven- or eleven-point rating scales to measure policy attitudes in web surveys.

1. Introduction

Web surveys have become an increasingly popular means of tracking the dynamics of policy attitudes and voting behavior during electoral campaigns (Ansolabehere and Rivers, 2013; Clarke et al., 2008). Despite concerns about coverage errors and sample representativeness web surveys offer several advantages. In contrast to telephone and face-to-face interviews, data collection is quick and inexpensive, which allows rapidly changing policy attitudes to be detected with large sample sizes (Berrens et al., 2003; Sanders et al., 2007; Vavreck and Rivers, 2008). In addition, web surveys facilitate the use of complex visual and auditory stimuli to study the effects of campaign activities on attitude formation and change (Vavreck and Iyengar, 2011).

A major challenge in conducting electoral research on the internet is determining how to measure attitudes toward contentious political issues. Traditionally, policy attitudes have been measured using seven- or eleven-point rating scales with endpoints that describe opposing political views on an issue. These scale formats have been criticized for being overly complex and ambiguous and for over-burdening respondents' cognitive abilities (Aldrich et al., 1982; Krosnick and Berent, 1993). According to satisficing theory (Krosnick, 1991, 1999), the quality of a survey response depends on the complexity of the judgment task at hand. When survey respondents find it difficult to understand questions about issues, they may be particularly inclined to use heuristic response strategies, which compromises the measurement quality of the reported policy attitudes. Some scholars have argued that a promising solution to this problem is to use branched rating scales (Dillman, 2008; Malhotra et al., 2009; Steeh, 2008) that divide the attitude-measurement process into two steps. The first step asks respondents about the direction of their attitudes, whereas the second step measures the intensity of respondents' attitudes. Past research suggests that reducing task complexity through branching formats indeed increases attitude stability and the relationship between issue attitudes and political behavior (Albaum, 1997; Aldrich et al., 1982; Krosnick and Berent, 1993). Consequently, branched issue scales are widely used in prominent election surveys, such as the American National Election Studies (ANES).

However, previous research on this topic has relied almost exclusively on data from interviewer-administered surveys. Because web surveys differ from telephone and face-to-face surveys in many respects (Christian et al., 2007), it is important to understand how different question formats affect the measurement quality of policy attitudes in self-administered surveys on the internet. While many studies have investigated the visual design of different question formats in web surveys (Christian et al., 2007; Christian et al., 2009; Cook et al., 2001; Tourangeau et al., 2007), no one has attempted to assess the validity of branched and unbranched rating scales. In the present article, I address this gap by comparing branched rating scales against conventional seven- and eleven-point rating scales in terms of response distribution, response quality and association with voting choice. More broadly, the aim of the present study is to advise survey designers in their choice of different rating scale formats for political attitudes in web surveys.

My analysis of data from a unique survey experiment embedded in the 2009 German Longitudinal Election Study (GLES) shows that

different scale formats lead to only minor differences in response distributions across a range of policy issues. Furthermore, none of the three scale formats consistently performs better with respect to different indicators of satisficing (non-opinion responses and non-differentiation). Finally, as the analysis of voting behavior demonstrates, the impact of controversial issues does not vary with the scale format in multivariate models of party choice. Overall, these results call into question the general usefulness of branched issue scales for web surveys on political attitudes. In the next section, I briefly review the literature on scale format and task complexity. Then, I lay out how web surveys might influence the operation of branching scales, and in the sections that follow, I describe the survey data, the experimental design, and the main results of the analyses. Finally, I discuss the results with regard to practical implications for designers of electoral surveys.

2. Theoretical background

2.1. Rating scale formats and task complexity

In mass surveys, attitudes toward political issues are typically measured with rating scales in which respondents are asked to rate their own position and the perceived positions of candidates or parties on a scale from 1 to 7 or from 1 to 11 with labeled endpoints. This practice can be illustrated using the following example from the German Longitudinal Election Study (GLES): "Some people would like to see lower taxes even if that means some reduction in health, education and social benefits; others would like to see more government spending on health, education and social benefits even if it means some increases in taxes. In your view, what is the position of the political parties on this issue?" In face-to-face interviews, the scale is typically presented along with a showcard on which the individual scale points are numbered and the end points labeled with "lower taxes, even if it means fewer social welfare benefits " and "more social welfare benefits, even if it means higher taxes and charges".

The dominant theoretical perspective for explaining the level of cognitive effort respondents allocate when answering issue questions is Krosnick's (1991) satisficing theory, which distinguishes between an optimal and a heuristic response mode. In an optimal mode, respondents must proceed through a series of fairly challenging steps when they answer attitude questions (Schwarz and Bohner, 2001; Tourangeau et al., 2000). First, they must interpret the question to determine to which attitude object the question refers and which dimensions of judgment are relevant. After that they must retrieve the appropriate information from memory and integrate it into an evaluative judgment. Finally, they must fit this judgment into a given response format. Due to the complexity of political issues it seems likely that not all respondents devote the necessary cognitive effort to provide an optimal response. Instead of thinking carefully about their attitude with respect to a particular topic, they might try to shortcut the answer process using simple heuristics. Survey research has identified a number of heuristic response strategies that are used by respondents (Baumgartner and Steenkamp, 2001; Tourangeau et al., 2000), such as acquiescence, midpoint responses, non-differentiation, avoiding certain response categories (such as by using "don't know" and "no opinion" responses), and providing socially desirable or even haphazard responses. Ultimately, heuristic responding introduces (non-)random measurement errors, which can either inflate or deflate the correlations between variables (Baumgartner and Steenkamp, 2001).

Satisficing theory holds that cognitive effort is a function of the motivation and mental abilities of the respondents as well as the difficulty of the judgment task. When respondents are motivated and equipped with the necessary cognitive resources, they tend to provide optimal answers by searching their memories for question-relevant information and integrating that information into an overall judgment of the policy presented. When respondents are unwilling or unable to elaborate upon the question, they rely on simple shortcuts. Consistent

with these hypotheses, previous studies show that respondents with low education levels or those who have little interest in politics are more prone to employ response heuristics (Banducci and Stevens, 2015; Holbrook et al., 2007; Kaminska et al., 2010; Krosnick, 1999; Krosnick et al., 1996; Narayan and Krosnick, 1996). More importantly, the model also predicts that satisficing is more likely when answering survey questions is difficult. Question difficulty, as indicated by linguistic complexity, was found to encourage response order effects (Holbrook et al., 2007) and non-substantive responses (Lenzner, 2012).

An additional aspect that affects task difficulty is the format of the question. The basic idea of the branching format is that decomposing attitude judgments into direction and intensity components facilitates the processing of attitude-relevant information and thus reduces the complexity of the question answering process (Armstrong et al., 1975; MacGregor et al., 1988). In an early study, Aldrich et al. (1982) examined the effects of rating scale format on the reliability and validity of self and candidate placement on political issues. Using a two-wave panel survey, these authors found that fully labeled and branched issue scales generated more extreme responses and were better at predicting candidate preferences than traditional seven-point scales with labeled endpoints. Further evidence of the superiority of branched scales was reported by Krosnick and Berent (1993). In a series of studies, these authors explored how question formats were related to the stability of policy preferences and ideology over time and their relationships with other political attitudes. They showed that the test-retest reliability of policy attitudes increases when they are measured using fully labeled branching questions. Moreover, these questions also correlated more closely with other issues, presidential performance and party identification, which speaks for the higher criterion validity of branching questions. In addition, Albaum (1997) reports that branched Likert scales of economic attitudes generated more extreme responses and were a better predictor of product preferences than unbranched versions (see also Yu et al., 2003).

Taken together, previous studies have shown that the question format substantively influences the response distributions of attitude scales. Branched rating scales consistently lead to more extreme attitudes than unbranched rating scales. However, the evidence that branched rating scales outperform traditional issues scales in reliability and validity is not particularly convincing. These past studies have considerably extended our knowledge of scale format effects, but several limitations in methodology render certain interpretations of these findings problematic. First, most studies compared fully labeled, branched scales with partially labeled unbranched scales. As labeling affects the reliability of rating scales (Alwin, 2010), the superior performance of the branched scales might be the result of labeling effects, question format, or both. Second, with certain exceptions, manipulations of scale format have been confounded with time of measurement. Thus, it is unclear whether events that occur between subsequent waves of a study (e.g., media coverage of issues or campaign events) or the question format itself were responsible for the better reliability and validity of the branched scales. Third, previous work is implicitly based on the premise that branching reduces the tendency to satisfice by decreasing task complexity and respondent burden, which in turn enhance the reliability of branched attitude scales. However, the effect of branching on satisficing behavior has never been tested. An alternative explanation would be that branching does not increase reliability, but the variance of scale scores by generating more extreme responses (Krosnick and Berent, 1993). It is well known that greater variability of scale scores inflate correlation coefficients which are often used to assess the criterion validity of attitude measures (Nunnally and Bernstein, 1994).

More recently, a growing number of studies have also voiced concerns about the superior performance of branched attitude scales. Based on data from the 2000 American Election Study, Treier and Hillygus (2009) showed that a self-reported political ideology scale with labeled scale points was more closely related to economic and social policy