Research Article

Self-coding: A method to assess semantic validity and bias when coding open-ended responses

Rebecca A. Glazier[1], Amber E. Boydstun[2] (iD)
and Jessica T. Feezell[3]

Abstract

Open-ended survey questions can provide researchers with nuanced and rich data, but content analysis is subject to misinterpretation and can introduce bias into subsequent analysis. We present a simple method to improve the semantic validity of a codebook and test for bias: a "self-coding" method where respondents first provide open-ended responses and then self-code those responses into categories. We demonstrated this method by comparing respondents' self-coding to researcher-based coding using an established codebook. Our analysis showed significant disagreement between the codebook's assigned categorizations of responses and respondents' self-codes. Moreover, this technique uncovered instances where researcher-based coding disproportionately misrepresented the views of certain demographic groups. We propose using the self-coding method to iteratively improve codebooks, identify bad-faith respondents, and, perhaps, to replace researcher-based content analysis.

Keywords

Survey design, survey methodology, open-ended questions, content analysis

Introduction

Political scientists often use open-ended survey questions to allow respondents to express ideas in their own words, and then the researchers code those responses using a pre-established codebook. For example, a wealth of scholarship in American and comparative politics on policy and valence issue priorities relies on open-ended questions (e.g., by Gallup or MORI) that gauge citizens' perceptions of the most important problem/issue currently facing the country (for overviews see Green and Hobolt, 2008; Wlezien, 2005). Other studies use open-ended questions to assess citizens' ideological positions (Swyngedouw, 2001). And in the running example we use here, more than three decades of scholarship based on Iyengar's (1994) canonical study commonly use open-ended survey responses, often embedded in experiments, to examine how framing an issue differently can affect who/what people hold responsible for that issue. The approach of giving people open-ended survey questions and then having researchers code the responses can provide rich and nuanced information about respondents' attitudes, helping advance and inform these and other areas of research.

Yet, coding open-ended responses means inferring preferences and sentiments by applying summary codes, opening the possibility that some responses are coded out of line with respondents' intended statements. Relatedly, when we apply a standard codebook to all open-ended responses, we may end up with biased results, with some types of respondents more likely to have their responses miscoded—an error that has an impact on the semantic validity of the data. In short, by assuming that all respondents who mentioned a certain phrase meant the same thing and should be coded the same way, we run the risk of misrepresenting the true opinions of some (non-random) portion of our samples.

Here, we present a "self-coding" method where respondents first provide open-ended responses and then code their

[1]University of Arkansas, Little Rock, USA
[2]University of California, Davis, USA
[3]University of New Mexico, USA

Corresponding author:
Amber E. Boydstun, University of California, Davis, 469 Kerr Hall, One Shields Ave, Davis, CA 95616, USA.
Email: aboydstun@ucdavis.edu

own responses into categories.[1] We demonstrated this technique by comparing data from a traditional coding process using an established codebook to data using the self-coding method. We found that the response codes assigned according to each process were not within the acceptable range for standard intercoder reliability. Moreover, we found that the variation in agreement between the codebook-based coding and respondents' self-coding was significantly correlated with key demographic characteristics. This demonstration suggests that the standard approach of researchers coding open-ended responses may not only lead to misinterpretation of respondents' attitudes but also introduce bias into the data. However, by iteratively using the self-coding method, refining the codebook to account for problems identified, and then repeating until no more problems are uncovered, researchers can arrive at a codebook that is optimally positioned to measure the attitudes of the survey respondents. In addition, as online data quality concerns become more prevalent, self-coded open-ended questions could help ensure that respondents are acting in good faith. And, where appropriate, researchers might even choose to use the self-coding method for their entire data collection process in order to maximize validity while saving resources.

The benefits and challenges of open-ended survey questions

Public opinion research has long struggled with the trade-off between rich open-ended responses and easy-to-analyze closed-ended responses. This choice is particularly vexing for large datasets that require quantitative analysis, and it has meaningful consequences for the data collected (e.g., Reja et al., 2003; Schuman and Presser, 1996). Closed-ended questions are much easier to analyze, but suffer other limitations such as response priming and measurement error (Bishop, 1987; Krosnick and Alwin, 1987; Schuman and Presser, 1996; Schwarz et al., 1992).

Open-ended questions can provide responses that are specific and nuanced in ways that are useful to researchers seeking to understand public opinion (Hickey and Kipping, 1996; Lazarsfeld, 1944; Montgomery and Crittenden, 1977). Yet in coding open-ended statements into categories for analysis, researchers run the risk of misrepresenting respondents' true intentions. A noted example is research on the judicial political knowledge questions on the American National Election Study (Gibson and Caldeira, 2009), which are coded according to a strict codebook and most likely underestimate public knowledge of the court (Boudreau and Lupia, 2011).

Krippendorff addresses the concern over misinterpreting respondents' intentions in his discussion of "sematic validity," which he defines as "the extent to which the results of coding correspond to how ordinary readers, or better still the respondents themselves would categorize what they say" (Krippendorff, 2008). Semantic validity is especially salient when we consider the possibility that a researcher-designed codebook might systematically misinterpret the responses of those with specific demographic characteristics.

Self-coding as a diagnostic method for semantic validity

Accurately coding opinions into discrete categories is an arguably difficult coding task for a professional coder. Best practices call for developing a detailed codebook (i.e., a coding schema that instructs researchers to categorize one type of observation into category X, and another type of observation into category Y, etc.) based on a close read of a random sample of observations (e.g., Baumgartner et al., 2008; Simon and Xenos, 2000). Yet these approaches maximize the degree to which researchers can consistently code responses from the point of view of the codebook they designed. It is possible that codebooks could be improved even further—and in some cases even replaced—by asking respondents to code their own responses, since they should know best what they are trying to say.[2] Although survey research is fraught with various measurement error challenges, including those rising from the potential instability and/or malleability of attitudes, most researchers agree that respondents indeed have "true attitudes," and that our job is to try to capture them (Geer, 1988; Zaller and Feldman, 1992).[3]

We suggest that researchers embarking on quantitative content analysis of open-ended survey responses employ the "self-coding" diagnostic method to support the semantic validity of their codebook and guard against biased data. The self-coding method involves first asking respondents to offer an open-ended response to a question about their opinion on something, and then asking them to code their own open-ended responses into categories. Then, without consulting the respondents' self-codes, the researchers code the open-ended responses into the same categories using the codebook. Finally, researchers compare the self-codes to the researcher-based codes to assess their similarity overall and to test for any bias in the researcher-based coding, asking two research questions:

> RQ1: How similar is respondents' self-coding of their open-ended responses to the researcher-based coding of their responses using the codebook?

> RQ2: Do any variances in the degree of agreement between the self-coding and the researcher-based coding correlate with respondents' demographic characteristics?

Demonstration of the self-coding method

In October 2018, we recruited a sample of 1,746 respondents based in the United States using Amazon's Mechanical Turk (MTurk) platform (Berinsky et al., 2012; Coppock,

2019; Mullinix et al., 2016).[4] We recruited subjects with a brief call stating, "Complete an online survey about your attitudes and opinions, approximately 7 minutes" and offered to pay them $0.70 for a completed survey. We utilized best practices for MTurk data collection, including limiting respondents to those with an approval rating of 95% or higher and having completed 100 jobs or more (Peer et al., 2014). However, even with these measures, there are increasing concerns about data quality on MTurk (Ahler et al., 2021; Kennedy et al., 2020; Ryan, 2020), including inattentive respondents, foreign respondents, and spoofed IP addresses. As we will discuss, an added benefit of open-ended questions is their ability to help identify good-faith respondents (Ryan, 2020).

For other research-related purposes, we were interested in replicating and extending the foundational research by Iyengar (1994) on media framing effects and so we used his questions and codebook here.[5] Iyengar found that when respondents are exposed to news reports that frame poverty in an episodic way (stories about individual poor people), respondents are more likely to hold individuals (vs. government/society) responsible for the problem of poverty than when exposed to news reports that frame poverty in a thematic way (stories about broad trends in poverty rates). The study by Iyengar used open-ended responses that were then coded into 93 categories and later aggregated according to a codebook into three broader categories used for analysis: "Government/Society," "Individual," and "Other." Although this is only one codebook, this foundational study has led to 30 years of research furthering our understanding of how framing influences responsibility attribution for social problems, from gun violence (Haider-Markel and Joslyn, 2001) to business failures (Williams et al., 2011) to economic adversity and voter turnout (Arceneaux, 2003).

Following Iyengar, we asked our respondents to read a brief news report about poverty, randomized to include either an episodic or thematic frame, in line with the substantive interests of our larger research agenda (see Appendix A). We then asked, "In your opinion, what are the most important causes of poverty?" and allowed them to enter up to 10 open-ended responses. Unlike Iyengar's original survey, however, we added a follow-up question that piped in each cause of poverty the respondent had entered, surrounded by the following text:

> When you say that one of the most important causes of poverty is [response from text box], do you think that cause is mostly because of individual people and the choices they make or is it mostly because of how government and society are these days?
>
> - Mostly about government/society
> - Mostly about individual people
> - Mostly about something else

This question was repeated for each of the open-ended causes the respondent provided. Of the 1,746 respondents, 1,541 provided at least one cause of poverty (on average each respondent offered 3.8 causes). In total, respondents offered 6,649 unique, open-ended causes of poverty, which they then coded into one of the three categories listed above. We then coded a sample of the open-ended responses ourselves using Iyengar's codebook. Finally, we focused on four specific codes (Education, Laziness, Unemployment, and Cost of Living), and used a series of tests to compare the respondents' coding of their own responses to our researcher-based coding in order to assess the semantic validity of our codebook and identify any potential bias in our researcher-based coding.

Analysis

We performed our researcher-based coding by drawing a random sample of 200 respondents (approximately 13% of the total 1,541 respondents who offered at least one response), yielding 740 open-ended responses. Each of these open-ended responses was then coded by a researcher according to the three higher-level codes (Government/Society, Individual, and Other) using the original codebook. Crucially, this was done without looking at the self-codes the respondents gave to their open-ended responses. We then calculated the intercoder reliability between our traditional coding and the respondents' self-coding, dropping responses that were coded "Other" by either respondents or researchers, leaving 667 responses to compare.[6] Comparing the two sets of "coders"—the researcher and the survey respondents—the agreement was 75% but the Cohen's Kappa was only 0.46, indicating substantial coding divergence (Freelon, 2013). In other words, when we asked respondents to code themselves, they provided codes that differed substantially from what we assumed when following the codebook.

To better understand this disagreement, we looked closely at four categories of causes given for poverty: Education, Laziness, Unemployment, and Cost of Living. We selected these categories for two reasons. First, they offer balance on higher-order codes: Education and Laziness fall under Individual in the original codebook, whereas Unemployment and Cost of Living fall under Government/Society. Second, in our previous replication of Iyengar's study (Feezell et al., 2019), these four categories were some of the most frequently coded. In that study, 897 respondents gave a total of 2,699 responses on the causes of poverty, which researchers then coded into the same 93 possible coding categories mentioned above: 10% were coded as Education, 5% as Laziness, 16% as Unemployment, and 4% as Cost of Living.

In order to identify all the responses that should fall under each of these four categories, we developed a set of keywords to flag any responses that might be about the given category (see Appendix B for details). The responses identified by the keyword searches for each category were then checked by hand and false positives were eliminated.
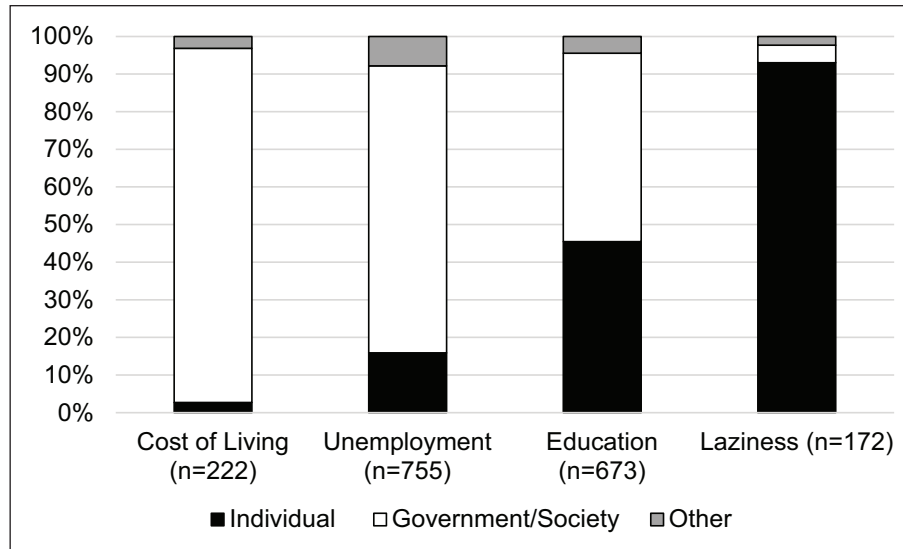
Figure 1. Respondents' Self-Coding of Four Causes of Poverty.

To clarify: our intent in this exercise was not to assign each response in our dataset a specific code, or even to examine all the instances of responses under these four causal categories according to the codebook. Rather, our goal was to select a subset of responses we were certain belonged in these four categories of the codebook as applied by researchers, allowing us to compare whether the self-codes for these items (Government/Society, Individual, Other) assigned by respondents matched the researcher codes assigned using the codebook.

Figure 1 shows the respondents' self-codes of their responses for each specific causal category (Cost of Living, Unemployment, Education, Laziness) into the three higher-level categories of Individual, Government/Society, and Other. When a respondent identified Cost of Living as a cause of poverty (the first column of Figure 1), the codebook would have instructed researchers to code this response as attributing responsibility to Government/Society. Respondents' self-codes reflected a similar instinct, with 93% of respondents self-coding their Cost of Living responses as Government/Society. In this case, the codebook and the respondents performed quite similarly on the coding task.

The second column of Figure 1, however, shows significant divergence. Seventy-six percent of respondents who identified an Unemployment response coded it as the responsibility of Government/Society (in line with our codebook), but nearly 16% of respondents coded their Unemployment responses as Individual instead. The vast majority of all the Unemployment responses were short phrases like "no jobs" or "unemployment," meaning researchers would be hard-pressed to find insights in the open-ended responses that would indicate to whom/what the respondent attributed responsibility.

The third column of Figure 1 shows responses that identify Education as a cause of poverty. Following the codebook directives, each Education response should be categorized as attributing responsibility to the Individual. When respondents coded themselves, however, they were almost evenly split between attributing responsibility to the Individual (45%) and Government/Society (50%). That is, half of the respondents indicated the researcher-assigned code was not an accurate interpretation of their response.

Finally, the fourth column of Figure 1 shows that the vast majority (94%) of Laziness causes were self-coded by respondents as Individual, consistent with the original codebook. Thus, for two of the topic categories in Figure 1, the codebook and the respondents performed similarly, but for two of them, there was substantial divergence. Looking both at Figure 1 and at the open-ended responses, we can see that while our strict codebook rules allowed researchers to easily categorize responses as either Government/Society or Individual (leaving only four Other responses), our respondents saw more nuance. They selected Other for 73 of the 740 responses, choosing to categorize responses like "lack of education," "lack of employment," and "bad luck" as Other.

This mismatch in terms of the codes assigned by the codebook and those that respondents assigned themselves should give researchers pause. When we assume what respondents mean by their responses and apply strict codebook rules, we risk misrepresenting respondent attitudes and biasing results in the initial coding and subsequent analysis.

Even more concerning is the possibility that we may be systematically biased in our interpretation of verbatim response coding. To assess this potential, we looked at the two categories with the greatest divergence in

Table 1. Testing demographic differences in attributing responsibility to Government/Society (vs. the Individual) for two identified causes of poverty: Education and Unemployment.

	Cause: Education			Cause: Unemployment		
	N	Mean	P value	N	Mean	P value
Gender						
Women	297	0.512	0.661	343	0.851	0.047
Men	335	0.528		288	0.802	
Race						
Non-white	120	0.458	0.935	126	0.877	0.052
White	512	0.535		506	0.817	
Education						
Some college, no degree	51	0.549	0.336	72	0.806	0.718
Associates degree or higher	581	0.518		560	0.832	
Income						
$75k or less	126	0.651	0.001	138	0.866	0.092
More than $75k	506	0.488		494	0.819	

coding among the four we examined: Education and Unemployment. We ran *t*-tests comparing the mean scores for the Government/Society attribution of responsibility for both categories, based on gender, race, education, and income. Given established findings that historically underprivileged groups tend to see more value in government support programs (e.g., Eagly et al., 2004; Hunt, 2007), we used one-tailed tests hypothesizing that respondents who are female, non-white, with lower education, and lower income will be more likely to see the problem of poverty as resting with Government/Society than their more privileged counterparts.

The rows in Table 1 offer comparisons of the extent to which different demographic groups attributed responsibility to Government/Society, out of the total amount of responsibility items they attributed to the Individual or to Government/Society (thus dropping "Other" items). Looking at the Education-related responses in the first column, we see significant differences by income; those who reported they made less than $75,000 a year (the bottom third of the sample) were also significantly more likely to attribute responsibility for Education-related causes to Government/Society ($p = 0.001$). If we were to code using the codebook, without asking respondents to first validate our coding, the results would be systematically biased against the accurate representation of the opinions of less wealthy respondents.

Turning to Unemployment-related responses in the second column, we see that women were significantly more likely than men to attribute responsibility for poverty as a result of Unemployment to Government/Society ($p = 0.047$). Non-white respondents were also more likely to attribute responsibility for poverty as a result of Unemployment to Government/Society, compared with white respondents ($p = 0.052$). The final set of comparisons in Table 1 indicate a likely difference based on income,

with less wealthy respondents more likely to attribute responsibility for poverty as a result of Unemployment to Government/Society ($p = 0.092$). Whereas in the case of Education the codebook appeared to be biased in a way that misrepresents the views of less wealthy respondents, in the case of Unemployment we found the views of women, white respondents, and less wealthy respondents were misrepresented.

But to what extent does traditional researcher-based coding harm our overall analysis? For our study of whether news framing of poverty as episodic or thematic affects respondents' responsibility attributions, we tried to replicate the substantive findings of Iyengar's original study. Specifically, we used respondents' self-coded data to calculate whether respondents in the thematic condition were more likely to attribute responsibility to Government/Society (vs. to the Individual), compared with respondents in the episodic condition. These analyses confirmed the substantive findings already documented in the literature: respondents who received the thematic news frames were more likely to provide open-ended responses on the causes of poverty that attributed responsibility to government and society ($p = 0.004$).

Yet the fact that we replicated the substantive findings of Iyengar's earlier research does not change our documented evidence that, for two major codes utilized, we are misrepresenting many respondents' intended answers, and in a non-random way. The self-coding method allows researchers to identify inconsistencies between self-codes and researcher-assigned codes and improve their codebooks accordingly. As our test case has illustrated, if drawing a new sample of respondents to engage in self-coding is a possibility, the self-coding method could be used in replication studies to ensure previous findings are robust to any demographic-related bias in how the codebook treated respondents' answers.

Applying the self-coding method

Researchers can apply the self-coding method in multiple ways. In the case of our demonstration above, we could refine the codebook instructions based on the problems related to semantic validity and bias we found, for example by coding responses involving education and unemployment into finer grained categories, some of which would be categorized as Individual and some as Government/Society.[7] Our test would then be repeated, iterating until we reach satisfactorily high levels of semantic validity and low levels of bias. Or at a minimum, if unable to change a codebook long-established in the literature, we would be able to interpret our substantive findings in the context of any bias we discovered through the self-coding method, potentially opening up new avenues of research. If at any point we were satisfied with the broad nature of the coding categories we gave respondents, we could decide to proceed with the "code yourself" method for the remainder of our data collection, perhaps repeating the researcher-based coding of a random sample of observations as a reliability check. Either way, we would have a stronger understanding of the degree to which our original conceptualization of categories matched—and did not match—how people themselves think about their responses.

As an added benefit, the open-ended questions used in the self-coding method are also valuable because they can help identify good-faith respondents on online survey platforms like MTurk. Professional survey-takers who are completing surveys as quickly as possible, or foreign survey-takers who do not understand the nuances of experimental treatments, are more likely to skip open-ended questions or to give nonsensical responses or responses that frequently repeat words like "good" and "nice" (Ryan, 2020). Researchers could search for these words, or create a word matrix to identify outlying words, to identify respondents they may want to exclude from analyses. High-level open-ended questions may also lead bad-faith actors to skip those questions altogether (Ryan, 2020), making it easy for researchers to identify them (approximately 12% of our original respondents did not provide any open-ended responses at all). The increasing use of web surveys has the potential to revitalize the use of open-ended questions (Smyth et al., 2009). By including open-ended questions, researchers can not only get richer data, but they can also quickly identify bad-faith respondents to be removed from the sample.

In addition, the self-coding method may serve as an effective way to measure individual-level attitudes without priming respondents with closed-ended options, leading to more accurate opinion measurement, of particular use when investigating questions with high degrees of social desirability. As one final advantage to this technique, contingent on agreement between self-coded and researcher-coded responses, the self-coding method may serve as a cost-effective way to accurately code large samples of open-ended responses.

Conclusion

Open-ended questions can give researchers a depth of understanding on complicated issues, like the causes of poverty. But sometimes we want to categorize those detailed responses into broad groupings. In these cases, a self-coding method could help ensure that researchers do so in a way that reflects the intent of the respondent.

We offer the self-coding method as one way to help researchers improve the semantic validity of open-ended response coding and help reduce unintentional bias introduced in the coding process. Our demonstration of this method revealed an unacceptably low level of intercoder reliability between our coding of responses according to an established codebook and to the respondents' own coding of their responses. Moreover, we showed cases where respondents' demographic factors were significantly correlated with their choice of code, meaning that if we coded their responses for a given cause into one lump code, we would be misrepresenting some demographic groups significantly more than others. By using the self-coding method, researchers could develop more nuanced codebooks or, were they using set codebooks for longitudinal or replication studies, they could at least be aware of the non-random ways in which the codebook might be misrepresenting respondents' intentions.

It is important to note that our test relied on a single codebook. Yet, other codebooks may suffer from similar problems of semantic validity manifested in different ways. It is easy enough to realize in hindsight that our codebooks should be nuanced. But as the 30-year-old literature based on—and continuing to use—Iyengar's responsibility attribution codebook illustrates, seeing the potential problems with a codebook is not always as easy or practical as we would want, especially given that researchers traditionally develop codebooks without any iterative "feedback" from the non-researcher humans whose latent attitudes they are designing these codebooks to capture. In cases where researchers are building on—and thus need to start by replicating—past research using established codebooks, researchers could use the self-coding method at least to be aware of any non-random ways in which the codebook might be misrepresenting respondents' intentions.

For example, with better insights from respondents, we could differentiate how *important* people think issues are from how much of a *problem* they think issues are when responding to Gallup's "Most Important Problem" question (Wlezien, 2005). Self-coding could also inform research on candidate evaluations, including whether respondents are considering policy or character/valence considerations in making their electoral choices (Gilens et al., 2007). The self-coding method we propose is not a silver bullet for the challenges of open-ended questions, but it could help researchers get closer to understanding what respondents really mean—by asking them.

Declaration of conflicting interests

Funding

ORCID iD

Amber E. Boydstun (iD) https://orcid.org/0000-0001-6012-8055

Supplementary materials

The supplementary files are available at http://journals.sagepub.com/doi/suppl/10.1177/20531680211031752.

Notes

1. See Roberts et al. (2014) for a computational modeling approach to the same problem we address here.
2. Note that there is no reason to expect the self-coding method to insulate researchers from the problem of social desirability influencing people's responses (Heerwegh, 2009; Krumpal, 2013).
3. But see also Schuman and Scott (1987) and Rahn et al. (1994).
4. Replication materials available on Harvard Dataverse at https://doi.org/10.7910/DVN/E4AJZF.
5. We are immeasurably grateful to Shanto Iyengar for sharing a copy of his original (hand-written!) codebook with us and encouraging our research in this area.
6. See Appendix C for details on self-coding frequencies and researcher intercoder reliability.
7. Another option would be to offer participants the option of a "Both" category, allowing participants to equivocate between the main closed-ended options being considered. If a high percentage of participants pick the Both category, it begs the question whether the closed-ended categories are nuanced enough.

Carnegie Corporation of New York Grant

References

Ahler DJ, Roush CE and Sood G (2021) The micro-task market for "lemons": Collecting data on Amazon's Mechanical Turk. *Political Science Research and Methods*. Forthcoming.

Arceneaux K (2003) The conditional impact of blame attribution on the relationship between economic adversity and turnout. *Political Research Quarterly* 56(1): 67–75.

Baumgartner FR, De Boef SL and Boydstun AE (2008) *The Decline of the Death Penalty and the Discovery of Innocence*. New York: Cambridge University Press.

Berinsky AJ, Huber GA and Lenz GS (2012) Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis* 20(3): 351–368.

Bishop GF (1987) Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly* 51(2): 220–232.

Boudreau C and Lupia A (2011) Political knowledge. In: Druckman JN, Green DP, Kuklinski JH and Lupia A (eds) *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press.

Coppock A (2019) Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Science Research and Methods* 7(3): 613–628.

Eagly AH, Diekman AB, Johannesen-Schmidt MC, et al. (2004) Gender gaps in sociopolitical attitudes: A social psychological analysis. *Journal of Personality and Social Psychology* 87(6): 796–816.

Feezell JT, Glazier RA and Boydstun AE (2019) Framing, identity, and responsibility: Do episodic vs. thematic framing effects vary by target population?. *Politics, Groups, and Identities* 9(2): 347–368.

Freelon D (2013) ReCal OIR: Ordinal, interval, and ratio intercoder reliability as a web service. *International Journal of Internet Science* 8(1): 10–16.

Geer JG (1988) What do open-ended questions measure? *Public Opinion Quarterly* 52(3): 365–367.

Gibson JL and Caldeira GA (2009) Knowing the Supreme Court? A reconsideration of public ignorance of the high court. *The Journal of Politics* 71(2): 429–441.

Gilens M, Vavreck L and Cohen M (2007) The mass media and the public's assessments of presidential candidates, 1952–2000. *The Journal of Politics* 69(4): 1160–1175.

Green J and Hobolt SB (2008) Owning the issue agenda: Party strategies and vote choices in British elections. *Electoral Studies* 27(3); 460–476.

Haider-Markel DP and Joslyn MR (2001) Gun policy, opinion, tragedy, and blame attribution: The conditional influence of issue frames. *Journal of Politics* 63(2): 520–543.

Heerwegh D (2009) Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research* 21(1): 111–121.

Hickey G and Kipping C (1996) A multi-stage approach to the coding of data from open-ended questions. *Nurse researcher* 4(1): 81–91.

Hunt MO (2007) African American, Hispanic, and white beliefs about black/white inequality, 1977-2004. *American Sociological Review* 72(3): 390–415.

Iyengar S (1994) *Is Anyone Responsible? How Television Frames Political Issues*. Chicago: University of Chicago Press.

Kennedy R, Clifford S, Burleigh T, et al. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods* 8(4): 614–629.

Krippendorff K (2008) Coding verbal data – What to optimize. In: *ANES conference on optimal coding of open-ended survey data*, 4–5 December 2008, Ann Arbor, MI: Institute for Social Research, University of Michigan. Available at: https://electionstudies.org/wp-content/uploads/2018/03/Krippendorff_transcript.pdf (accessed 12 July 2021).

Krosnick JA and Alwin DF (1987) An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly* 51(2): 201–219.

Krumpal I (2013) Determinants of social desirability bias in sensitive surveys: A literature review. *Quality and Quantity* 47(4): 2025–2047.

Lazarsfeld PF (1944) The controversy over detailed interviews—an offer for negotiation. *Public Opinion Quarterly* 8(1): 38–60.

Montgomery AC and Crittenden KS (1977) Improving coding reliability for open-ended questions. *Public Opinion Quarterly* 41(2): 235–243.

Mullinix KJ, Leeper TJ, Druckman JN, et al. (2016). The generalizability of survey experiments. *Journal of Experimental Political Science* 2(2): 109–138.

Peer E, Vosgerau J and Acquisti A (2014) Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods* 46(4): 1023–1031.

Rahn WM, Krosnick JA and Breuning M (1994) Rationalization and derivation processes in survey studies of political candidate evaluation. *American Journal of Political Science* 38(3): 582–600.

Reja U, Manfreda KL, Hlebec V, et al. (2003) Open-ended vs. close-ended questions in web questionnaires. *Developments in Applied Statistics* 19(1): 159–177.

Roberts ME, Stewart BM, Tingley D, et al. (2014) Structural topic models for open-ended survey responses. *American Journal of Political Science* 58(4): 1064–1082.

Ryan TJ (2020) Fraudulent responses on Amazon Mechanical Turk: A fresh cautionary tale. Available at: https://timryan.web.unc.edu/2020/12/22/fraudulent-responses-on-amazon-mechanical-turk-a-fresh-cautionary-tale/ (accessed 12 July 2021).

Schuman H and Presser S (1996) *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks: Sage Publications.

Schuman H and Scott J (1987) Problems in the use of survey questions to measure public opinion. *Science* 236(4804): 957–959.

Schwarz N, Hippler H-J and Noelle-Neumann E (1992) A cognitive model of response-order effects in survey measurement. In: Schwartz N and Sudman S (eds) *Context Effects in Social and Psychological Research*. New York: Springer, 187–201.

Simon A and Xenos M (2000) Media framing and effective public deliberation. *Political Communication* 17(4): 363 – 376.

Smyth JD, Dillman DA, Christian LM, et al. (2009) Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly* 73(2): 325–337.

Swyngedouw M (2001) The subjective cognitive and affective map of extreme right voters: Using open-ended questions in exit polls. *Electoral Studies* 20(2): 217–241.

Williams AE, Davidson R and Yochim EC (2011) Who's to blame when a business fails? How journalistic death metaphors influence responsibility attributions. *Journalism and Mass Communication Quarterly* 88(3): 541–561.

Wlezien C (2005) On the salience of political issues: The problem with 'most important problem'. *Electoral Studies* 24(4): 555–579.

Zaller J and Feldman S (1992) A simple theory of the survey response: Answering questions versus revealing preferences. *American Journal of Political Science* 36(3): 579–616.