

Much ado about acquiescence: The relative validity and reliability of construct-specific and agree–disagree questions

Research and Politics
July–September 2015: 1–8
© The Author(s) 2015
DOI: 10.1177/2053168015604173
rap.sagepub.com


Yphtach Lelkes¹ and Rebecca Weiss²

Abstract

Acquiescence response bias, or the tendency to agree with questions regardless of content, is a prominent concern in survey design. An often proposed solution, and one that was recently implemented in the American National Election Study, is to rewrite response options so that they tap directly into the dimensions of the construct of interest. However, there is little evidence that this solution improves data quality. We present a study in which we employ two waves of the 2012 American National Election Study in order to compare the reliability and concurrent validity of political efficacy questions in both the agree–disagree and construct-specific formats. Construct-specific questions were not only as reliable and valid as agree–disagree questions generally, they were also as valid among respondents that were most likely to acquiesce. This suggests two possible outcomes: Either agree–disagree questions do not negatively impact data quality or that construct-specific questions are not a panacea for acquiescence response bias.

Keywords

American National Election Study, acquiescence, surveys

Researchers have been aware of acquiescence response bias, i.e. the tendency to agree with questions regardless of their content, for at least half a century (Bass, 1955; Christie et al., 1958; Jackson and Messick, 1957; Peabody, 1961). While a number of suggestions to attenuate its impact have been proposed, the solution endorsed as best practice by a number of handbook chapters and review articles (Krosnick, 1989, 1999; Krosnick et al., 2005, 1996; Krosnick and Presser, 2010; Pasek and Krosnick, 2010; Schuman and Presser, 1996; Smyth et al., 2006; Vannette and Krosnick, 2014) suggests writing questions in a “construct-specific” manner, meaning that response options directly tap into the dimensions of the construct of interest. But to our knowledge, only one study explicitly compares the quality of data gathered from agree–disagree questions against construct-specific questions (Saris et al., 2010), and that study only evaluated the relative convergent validity and reliability of the two types of measures. Hence, we possess limited knowledge about the relative impact either question type has on the quality of gathered data.

In the following study, we employ the 2012 American National Election study (ANES) and compare the concurrent

validity of a set of agree–disagree questions against construct-specific alternatives. Not only do we not find differences between responses to question formats among the general population, but construct-specific questions are not better at attenuating acquiescence among respondents that are most likely to acquiesce. This indicates two possibilities: either the impact of acquiescence response bias on data quality is overstated or construct-specific alternatives do not address the problem of acquiescence response bias.

The problem of acquiescence response bias

Acquiescence response bias represents a significant threat to social science inference. Its systematic nature can alter the correlations between substantive variables of interest,

¹University of Amsterdam, the Netherlands

²Stanford University, USA

Corresponding author:

Yphtach Lelkes, University of Amsterdam, Nieuwe Achtergracht 166, Amsterdam, 1018 WV, the Netherlands.

Email: y.lelkes@uva.nl



Creative Commons Non Commercial CC-BY-NC: This article is distributed under the terms of the Creative Commons

Attribution-NonCommercial 3.0 License (<http://www.creativecommons.org/licenses/by-nc/3.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<http://www.uk.sagepub.com/aboutus/openaccess.htm>).

leading to a failure to accurately measure the size and strength of a relationship between variables, which can significantly impact the conclusions drawn by survey researchers (Alwin and Krosnick, 1991; Baumgartner and Steenkamp, 2001; Krosnick, 1999).

At least three mechanisms account for acquiescence response bias. First, some respondents may exhibit various personality traits that are associated with a tendency to agree (Adorno et al., 1950; Bass, 1955; Couch and Keniston, 1960; Edwards, 1961; Gage et al., 1957; Knowles and Condon, 1999; Samelson and Yates, 1967; Shaw, 1961) such as agreeableness or a tendency to conform (Krosnick and Fabrigar, 2015). Second, some people may acquiesce for social desirability reasons (Couch and Keniston, 1960; Knowles and Condon, 1999; Webb et al., 1981). As conversational conventions imply that the interviewer agrees with a statement, a respondent may say he or she agrees to garner a positive image. The third account, the satisficing perspective, draws from the cognitive basis of responding to survey questions. Respondents may first tend to think of reasons a claim is valid, and respondents who are not as motivated or are less cognitively capable may stop the response process before thinking of reasons why the statement may also be invalid (Knowles and Condon, 1999; Krosnick, 1991; Vannette and Krosnick, 2014).

A number of solutions to overcome the issues of acquiescence response bias have been suggested, including item reversals (Lichtenstein and Bryan, 1965). However, most primers on questionnaire design instead suggest “forced-choice” or “construct-specific” response scales (Krosnick, 1989, 1999; Krosnick et al., 2005, 1996; Krosnick and Presser, 2010; Pasek and Krosnick, 2010; Schuman and Presser, 1996; Smyth et al., 2006), wherein respondents are asked directly about the underlying dimension. For instance, rather than asking a respondent whether they agree or disagree with the statement “I approve of the way the president is handling his job,” a construct-specific question would instead ask a respondent how much they approve or disapprove of the way the president is handling his job. This suggestion is based on the notion of satisficing as construct-specific questions are less cognitively burdensome.

Based on this reasoning, construct-specific questions should be more valid than agree–disagree questions. But little empirical evidence directly assesses the validity of this conjecture. One study does indeed demonstrate that construct-specific response scales yield higher validity and reliability estimates than agree–disagree scales (Saris et al., 2010). These authors utilize multi-trait multi-method (MTMM) experiments in several countries and find that construct-specific questions yielded higher levels of convergent (or what the authors call “internal”) validity and are more reliable than agree–disagree questions.

While MTMM methods effectively assess convergent validity (Alwin and Krosnick, 1991; Saris et al., 2010),

validity is a multifaceted construct. Another aspect of validity, and one that is more likely to affect the conclusions we draw in social science research, is whether a target measure is related to another construct to which it is theoretically linked (Messick, 1989). Past public opinion research has adopted this form of validity, i.e. criterion validity, as indicative of measurement quality (e.g. Chang and Krosnick, 2009; Jenkinson et al., 1994; Malhotra and Krosnick, 2007; Parry and Crossley, 1950; Yeager and Krosnick, 2010). In this formulation, two forms of a survey question are randomly administered to different groups of respondents. One form is deemed to be more valid if it is more closely related to theoretically-linked criterion than the other. We adopt this technique in our current investigation.

The present study

Methods. Data came from the ANES 2012 Time Series study. The two-wave pre- and post-election study combines online and face-to-face interviews, and yielded a total sample of 5,916 interviews in the pre-election wave (2,056 face-to-face and 3,860 online) and 5,513 interviews in the post-election wave. The question batteries used to construct the target measures, along with other questions used to construct the criterion measures, can be found in the online appendix.¹

Target measures. Respondents were randomly assigned to receive either the agree–disagree form or the construct-specific form of a set of four items designed to measure political efficacy. Respondents that saw one set of items in the pre-election wave also saw the same set of items in the post election wave. These target measures and all criterion measures were rescaled to lie between 0 and 1, with 0 indicating the lowest possible value and 1 the highest. We also computed two indices, one that averages the internal efficacy measures (items 1 and 2) and one that averages the external efficacy measures (items 3 and 4), in order to reduce measurement error.

Criterion measures. We chose criterion items that have been shown to be correlated with the core constructs (internal efficacy and external efficacy) but do not share the response scale of the of either set of target items. This was to ensure that criterion and target item correlations were not artifacts of a similar response scale. Three indicators, all relating to political activism (the outcome measure in most political efficacy studies), met these requirements: (1) R’s level of political knowledge; (2) percent chance of voting; (3) political activism defined as the sum of items checked from a ten-item checklist. Following past studies using similar measures, we computed the number of items indicating the activities the respondents said he or she had done.

In addition to general population differences in the validity and the reliability of these two sets of questions, we

were able to test whether construct-specific questions would be particularly more valid among those that were likely to acquiesce. First, in line with satisficing perspective, we tested whether the response form would have a greater impact on validity among those with low versus high levels of verbal ability.² Addressing the personality account, we examined whether construct-specific questions yielded more valid data among respondents displaying high levels of agreeableness, which was measured via the agreeableness dimension on the ten-item personality inventory (Gosling et al., 2003). Finally, the social desirability perspective of acquiescence response bias leads us to hypothesize that the differences between construct-specific and agree–disagree questions would be the larger when interviews were conducted face-to-face rather than online (where social desirability pressures are lower).

Results

The agree–disagree questions and the construct-specific questions displayed similar levels of reliability. The test-retest polychoric correlations between each efficacy question asked in wave 1 and its counterpart in wave 2 were larger in four of the six cases for the agree/disagree questions than they were for the construct-specific questions (Table 1). However, the differences in test–retest correlations were not substantive. The average correlation of the pre- and post-wave measures for the agree/disagree questions was $r = 0.63$, while the average correlation for the pre- and post-wave measures for the construct-specific version was also $r = 0.63$.

Figure 1 displays the unstandardized coefficients and 90% confidence intervals from bivariate models predicting the criterion measures from each target measure.³ We present 90% confidence intervals as the robustness of the differences are more easily assessed (compared with p -values), and are a better indicator of “negligible differences” (Rainey, 2014). We plot the coefficients for ease of consumption (Kastellec and Leoni, 2007), but tables appear in the online appendix.

Each column within a panel displays the b coefficient from each of the six criterion questions. As the questions were asked twice, we display coefficients from models predicting the criterion between the criterion and the item when asked in Wave 1 (top row) and Wave 2 (bottom row). For instance, the top row of the top left facet of Panel A shows the coefficients and 90% confidence intervals from a model predicting the political knowledge from the construct-specific version of the first efficacy measure; the second row shows the coefficients from a model predicting the political knowledge from agree–disagree version of the second efficacy measure. Each panel displays a different criterion variable: Panel A displays the relationship between the efficacy measures and R’s political knowledge, Panel B shows the relationship between the efficacy

Table 1. Test–retest reliability estimates for agree–disagree item and construct-specific items.

	Standard	Revised	Difference
Efficacy 1	0.66	0.57	0.08
Efficacy 2	0.63	0.73	–0.09
Efficacy 3	0.58	0.55	0.03
Efficacy 4	0.61	0.60	0.01
External efficacy	0.63	0.61	0.02
Internal efficacy	0.69	0.72	–0.02

measures and the percent chance of voting, and Panel C shows the relationship between efficacy measures and R’s political activity.

The construct-specific and agree–disagree questions displayed indistinguishable levels of criterion validity in almost every instance. While the point estimates for the construct-specific questions predicting the respondents’ political knowledge was larger than the point estimates for the agree–disagree questions 7 out of 12 times (Panel A, Figure 1), the average difference between the criterion validity of the agree–disagree questions and the construct-specific questions was -0.02 ($\sigma = 0.03$), the 90% confidence overlapped in all but two instances. In one of these instances (Efficacy 4, Wave 2), the agree–disagree questions were more predictive of political knowledge than the construct-specific questions. Similarly, the point estimates of the coefficients predicting R’s political activity again reveal no substantive differences between the two question forms. This time the construct-specific form was larger than the agree–disagree form in all but two cases, but the average difference was only -0.05 ($\sigma = 0.07$). The 90% confidence intervals overlapped in six of the 12 cases. However, in one case (Efficacy 2, Wave 1), the agree–disagree question was more valid than the construct specific question.

Finally, the point estimates of the coefficients predicting R’s political activity were larger in every case, but the differences were quite small: the average difference between the criterion validity of the agree–disagree questions and the construct specific questions was -0.03 ($\sigma = 0.02$). The 90% confidence intervals overlapped in all but one case.

Next we tested whether construct-specific questions would be more valid than agree–disagree questions among people that we expect to be more susceptible to acquiescence response bias or in situations that increase acquiescence. We regressed each target measure on the respondents efficacy score, a dummy indicating whether they received the construct-specific or the agree–disagree form, and one of proposed moderators (their verbal ability scores, their agreeableness scores, or the mode of interview), the three-way interaction between these three variables, and the component two-way interactions. For the sake of presentation, we plot the three-way interactions from these models in

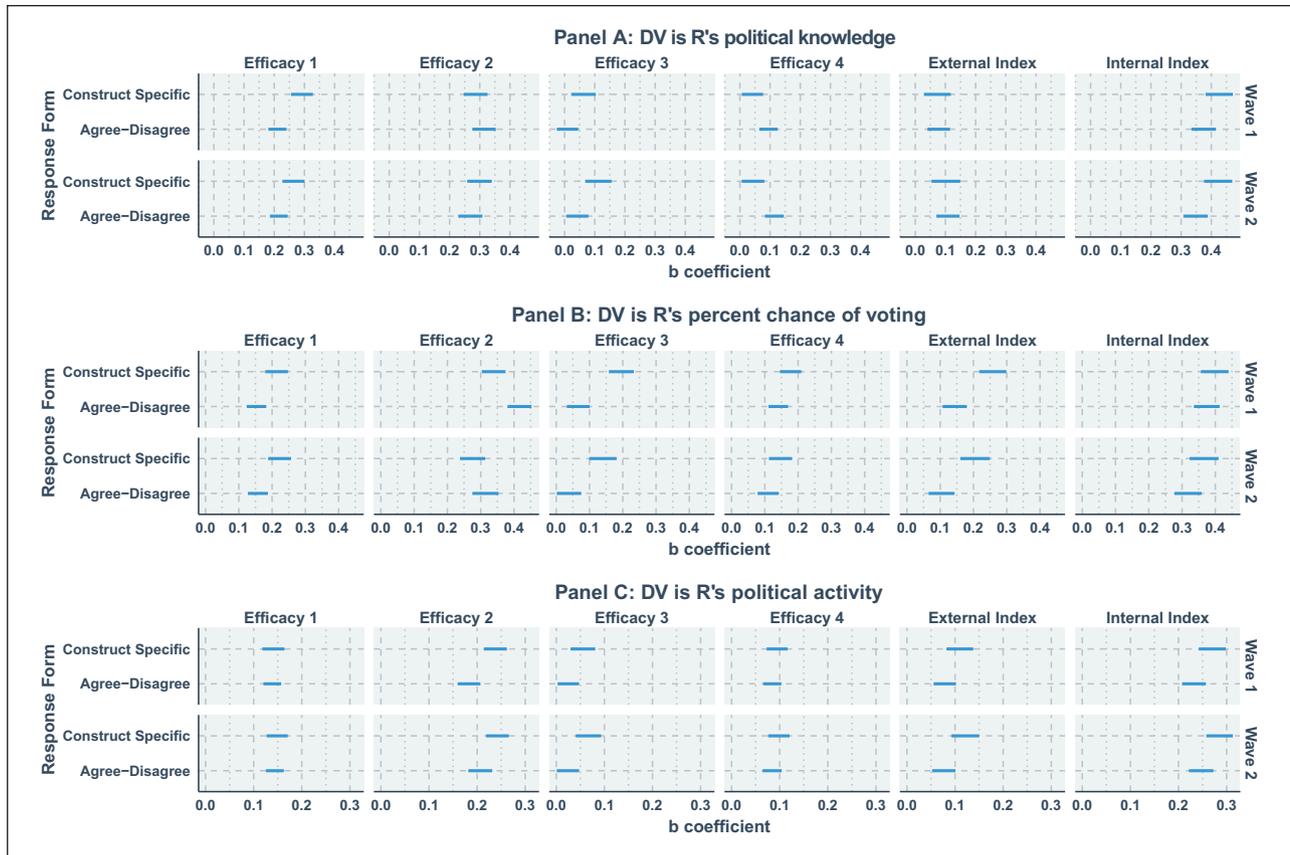


Figure 1. Relative validity of agree–disagree and construct specific efficacy questions.

Figures 2, 3, and 4, and include the tables in the online appendix.

Counter the cognitive burden hypothesis, construct-specific questions were no more valid than agree–disagree questions among those low in verbal ability than among those high in verbal ability. In Panel A of Figure 2, only two of the 12 coefficients from models predicting political knowledge did not overlap with zero. However, these coefficients do not align with our expectations. The construct-specific questions are more valid among those with higher cognitive ability scores. In the models predicting R's percent chance of voting, we again see that only 2 of the 12 coefficients do not overlap with zero: this time the coefficients are in the expected direction, construct-specific questions are more valid than agree–disagree questions among those with lower cognitive ability scores. Finally, cognitive ability did not moderate the differences between construct-specific and agree–disagree questions in models predicting political activity.

Similarly, we do not find any support for the hypothesis that construct-specific questions would be more valid among those high in agreeableness (Figure 3). One of the 12 coefficients from models predicting political knowledge did not overlap with zero. However, the direction of the effect was counter to our hypothesis: construct-specific

questions were more valid among those low in agreeableness. All of the three-way interactions from models predicting R's percent chance of voting overlapped with zero. Three of twelve coefficients from models predicting political activism did not overlap with zero, but, in a way that was counter to our expectation: Construct-specific questions were more valid among those low in agreeableness.

Finally, the construct-specific forms did not prove to be more valid in face-to-face interviews. The target measures were regressed on the respondents efficacy score, the survey mode, the response form, and the two-way and three-way interactions of interest (see Figure 4). The three-way interaction coefficients from models predicting political knowledge all overlapped with zero (Panel A). Nine of twelve three-way interactions predicting R's percent chance of voting overlapped with zero (Panel B). The coefficients that do not overlap with zero are consistent with our expectations: the construct-specific questions improve validity in the face-to-face interviews but not the Internet interviews. However, the three coefficients that do not overlap with zero from the next set of models (predicting political activity) show the opposite. The construct-specific questions yield more valid responses when surveys are conducted over the Internet than when they are conducted face-to-face.

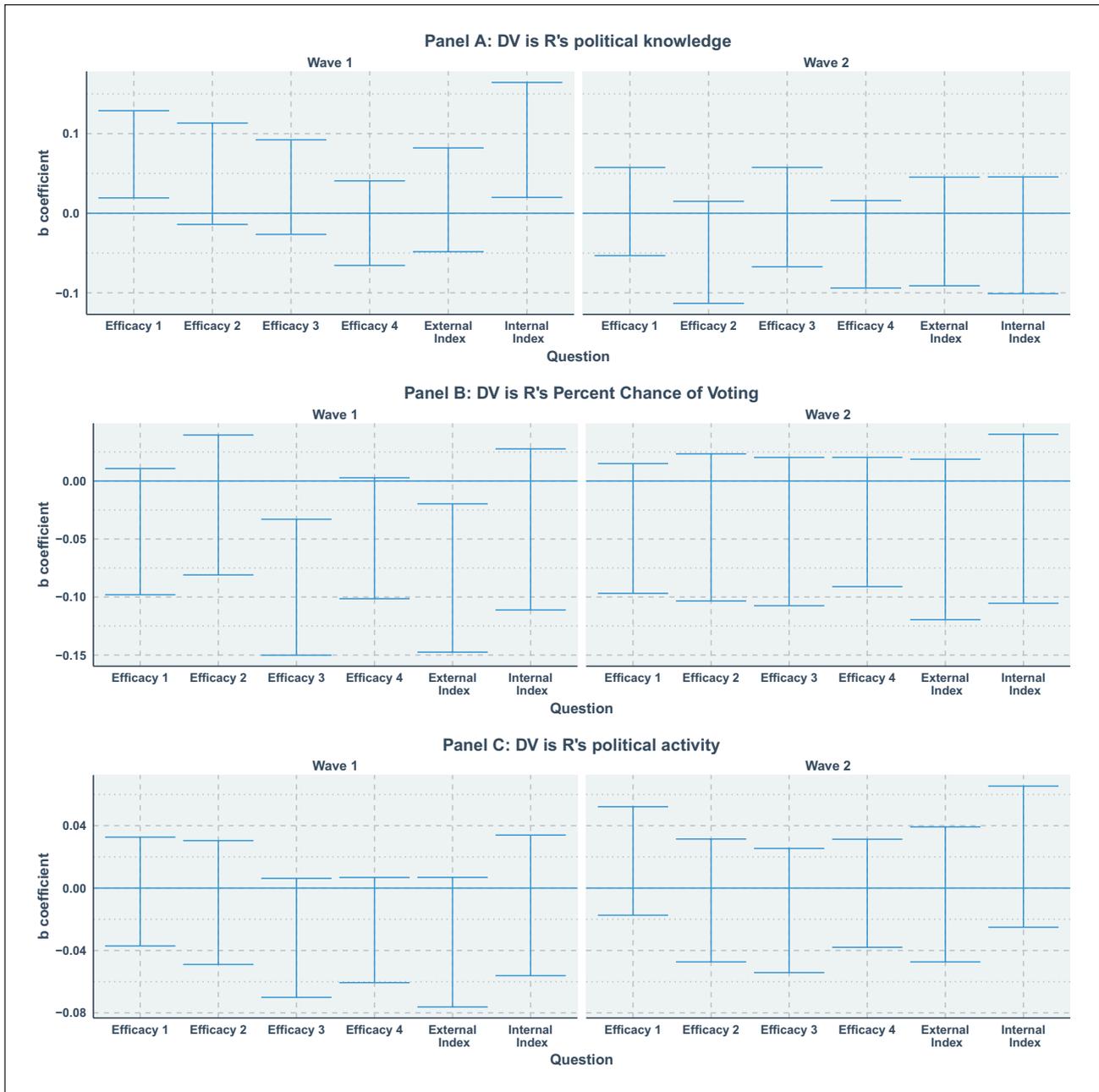


Figure 2. Three-way interaction coefficients from models testing **verbal ability** * response form * efficacy.

Discussion

Despite the compelling psychological theory behind the use of construct-specific questions as a solution to acquiescence response bias, little evidence shows that these questions actually yield better data quality than agree–disagree questions. Our results indicate that the reliability and concurrent validity of these optimized questions was no better than the agree–disagree questions. Even among those that were most likely to acquiesce (those with low verbal ability scores or those high in agreeableness) or those interviewed

face-to-face, construct-specific and agree–disagree questions were equally valid. This indicates that the results obtained through “optimized” questions are most likely equivalent to those obtained through agree–disagree questions.

It should be noted that the point estimates of the regression coefficients from the construct-specific regressions were consistently, albeit mostly trivially, larger than the agree–disagree questions. It could be argued that we might see significant differences with a larger sample. The size of the 2012 ANES, however, is already much

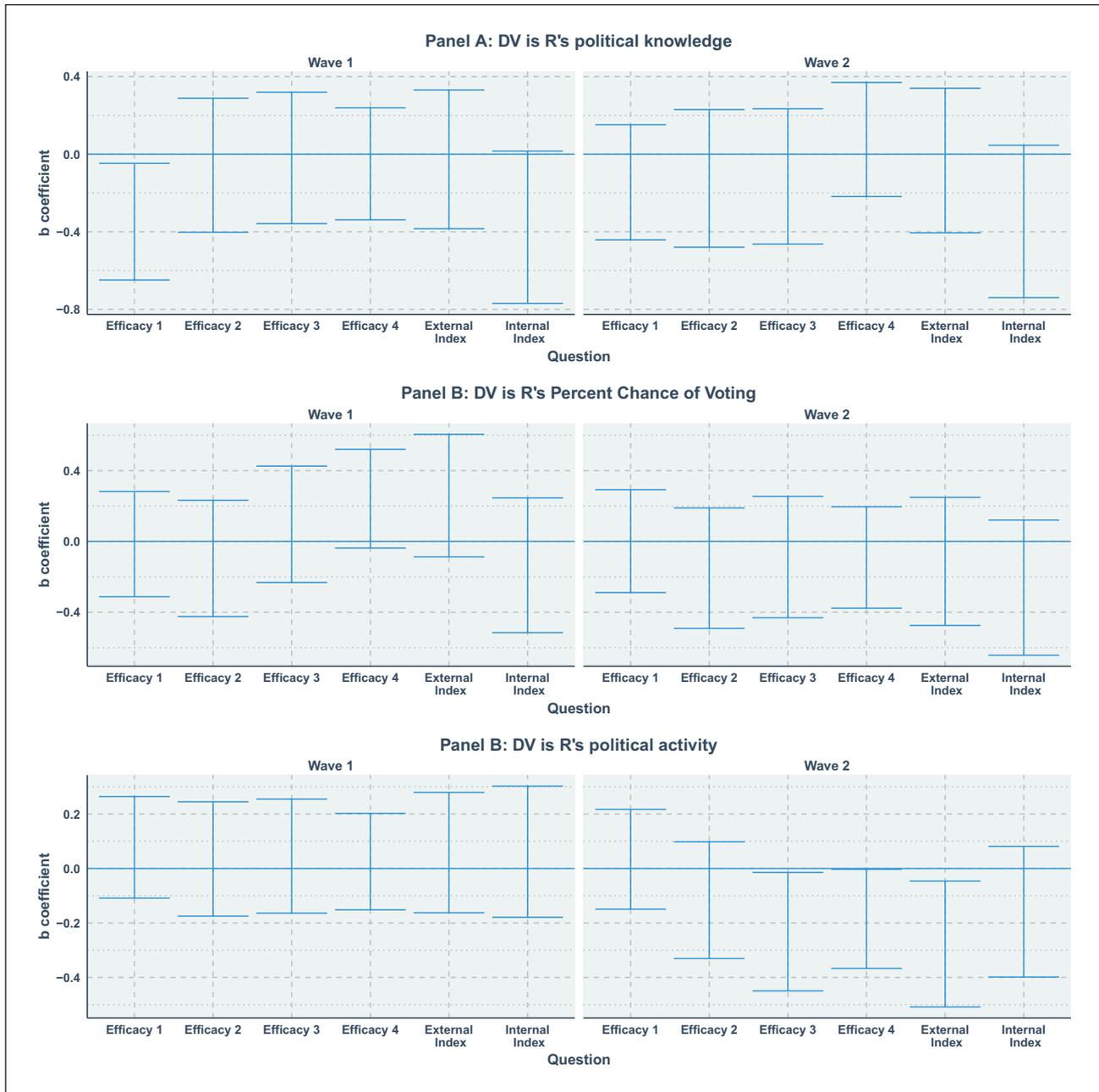


Figure 3. Three-way interaction coefficients from models testing **agreeableness** * response form * efficacy.

larger than most surveys, and differences would probably not be substantive.

Why were the construct-specific questions not more valid? We see three possibilities. First, acquiescence response bias may not actually affect the validity of survey data. Given abundant evidence of bias when agree-disagree questions are used, we are skeptical of this possibility. It may be the case, however, that the effect of acquiescence response bias on data quality is actually minimal.

Second, construct-specific questions are not actually optimal, at least in terms of easing acquiescence response

bias. It could certainly be the case that the cognitive burden of agree-disagree questions or the cognitive ease of construct-specific questions has been overstated. Construct-specific questions are tailored to ease cognitive burden (see Vannette and Krosnick, 2014), but if acquiescence is actually caused by another factor, e.g. motivational purposes, then we require an alternative solution.

Finally, it is possible that, generally speaking, construct-specific questions provide better quality data than agree-disagree questions, but not in this specific case we present here. This experiment was limited to questions

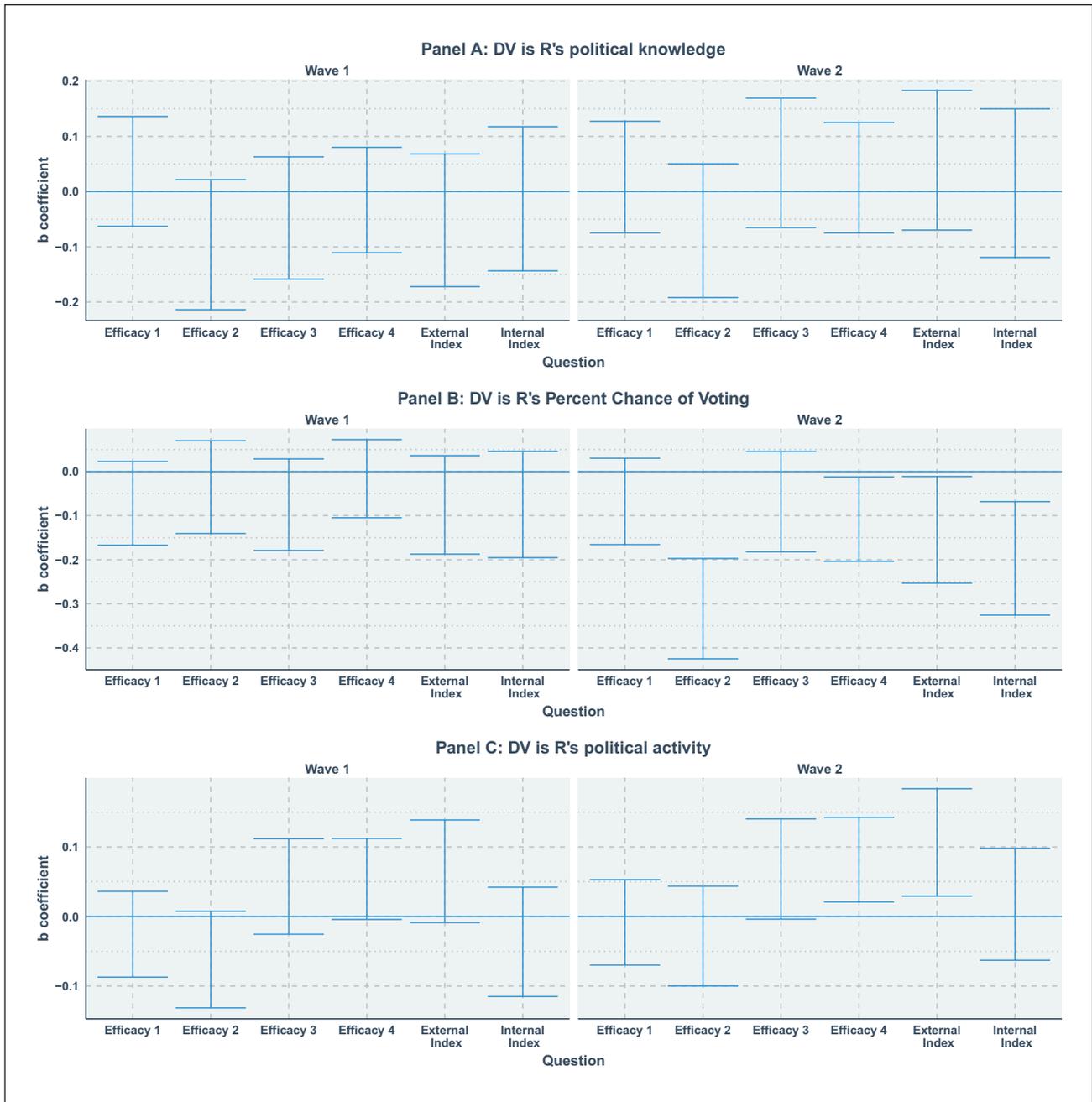


Figure 4. Three-way interaction coefficients from models testing **survey mode** * **response form** * **efficacy**.

pertaining to political efficacy. Future studies should expand the domain of questions. Since the construct-specific forms were not more valid in one instance, our results suggest that construct-specific formats should not be considered a canonical solution to the problem of acquiescence response bias.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. Detailed information about the sample can be found at <http://www.electionstudies.org>.
2. The ANES included the Wordsum vocabulary test. We fit respondents' responses to this test to a two-parameter logistic IRT model (Cor et al., 2012), and extract individual ability estimates.
3. For the sake of consistency with past research, we treat each efficacy measure as continuous. Results are substantively the same if we instead utilize polyserial or polychoric correlations.

Supplementary material

The online appendix is available at: <http://rap.sagepub.com/content/by/supplemental-data>

References

- Adorno TW, Frenkel-Brunswick E, Levinson DJ and Sanford RN (1950) *The authoritarian personality*. Harpers.
- Alwin DF and Krosnick JA (1991) The reliability of survey attitude measurement: the influence of question and respondent attributes. *Sociological Methods and Research* 20(1): 139–181.
- Bass BM (1955) Authoritarianism or acquiescence? *The Journal of Abnormal and Social Psychology* 51(3): 616.
- Baumgartner H and Steenkamp JBE (2001) Response styles in marketing research: a cross-national investigation. *Journal of Marketing Research* 38(2): 143–156.
- Chang L and Krosnick JA (2009) National surveys via rdd telephone interviewing versus the internet comparing sample representativeness and response quality. *Public Opinion Quarterly* 73(4): 641–678.
- Christie R, Havel J and Seidenberg B (1958) Is the F scale irreversible? *The Journal of Abnormal and Social Psychology* 56(2): 143.
- Cor MK, Haertel E, Krosnick JA and Malhotra N (2012) Improving ability measurement in surveys by following the principles of IRT: The wordsum vocabulary test in the general social survey. *Social Science Research* 41(5): 1003–1016.
- Couch A and Keniston K (1960) Yeasayers and naysayers: Agreeing response set as a personality variable. *The Journal of Abnormal and Social Psychology* 60(2): 151.
- Edwards AL (1961) Social desirability or acquiescence in the MMPI? A case study with the SD scale. *The Journal of Abnormal and Social Psychology* 63(2): 351.
- Gage N, Leavitt GS and Stone GC (1957) The psychological meaning of acquiescence set for authoritarianism. *The Journal of Abnormal and Social Psychology* 55(1): 98.
- Gosling SD, Rentfrow PJ and Swann WB (2003) A very brief measure of the big-five personality domains. *Journal of Research in Personality* 37(6): 504–528.
- Jackson DN and Messick SJ (1957) A note on “ethnocentrism” and acquiescent response sets. *The Journal of Abnormal and Social Psychology* 54(1): 132.
- Jenkinson C, Wright L and Coulter A (1994) Criterion validity and reliability of the SF-36 in a population sample. *Quality of Life Research* 3(1): 7–12.
- Kastellec JP and Leoni EL (2007) Using graphs instead of tables in political science. *Perspectives on Politics* 5(04): 755–771.
- Knowles ES and Condon CA (1999) Why people say “yes”: A dual-process theory of acquiescence. *Journal of Personality and Social Psychology* 77(2): 379.
- Krosnick JA (1989) Attitude importance and attitude accessibility. *Personality and Social Psychology Bulletin* 15(3): 297–308.
- Krosnick JA (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5(3): 213–236.
- Krosnick JA (1999) Survey research. *Annual Review of Psychology* 50(1): 537–567.
- Krosnick JA and Fabrigar LR (2015) *Designing good questionnaires: Insights from psychology*. New York: Oxford University Press.
- Krosnick JA, Judd CM and Wittenbrink B (2005) The measurement of attitudes. In Albarracín D, Johnson BT and Zanna MP (eds), *The handbook of attitudes*. Mahwah, NJ: Erlbaum, pp. 21–76.
- Krosnick JA, Narayan S and Smith WR (1996) Satisficing in surveys: Initial evidence. *New Directions for Program Evaluation* 70: 29–44.
- Krosnick JA and Presser S (2010) Question and questionnaire design. In: Marsden PV and Wright JD (eds), *Handbook of survey research*, 2nd edn. Bradford: Emerald Group Publishing Limited, pp. 263–313.
- Lichtenstein E and Bryan JH (1965) Acquiescence and the MMPI: An item reversal approach. *Journal of Abnormal Psychology* 70(4): 290.
- Malhotra N and Krosnick JA (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.
- Messick S (1989) Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher* 18(2): 5–11.
- Parry HJ and Crossley HM (1950) Validity of responses to survey questions. *Public Opinion Quarterly* 14(1): 61–80.
- Pasek J and Krosnick JA (2010) Optimizing survey questionnaire design in political science: insights from psychology. *Oxford handbook of American elections and political behavior*. Oxford: Oxford University Press, pp. 27–50.
- Peabody D (1961) Attitude content and agreement set in scales of authoritarianism, dogmatism, anti-semitism, and economic conservatism. *The Journal of Abnormal and Social Psychology* 63(1): 1.
- Rainey C (2014) Arguing for a negligible effect. *American Journal of Political Science* 58(4): 1083–1091.
- Samelson F and Yates JF (1967) Acquiescence and the F scale. *Psychological Bulletin* 68(2): 91.
- Saris WE, Revilla M, Krosnick JA and Shaeffer EM (2010) Comparing questions with agree/disagree response options to questions with item-specific response options. *Journal of the European Survey Research Association* 4(1): 61–79.
- Schuman H and Presser S (1996) *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context (Quantitative Studies in Social Relations)*. London: SAGE Publications.
- Shaw ME (1961) Some correlates of social acquiescence. *The Journal of Social Psychology* 55(1): 133–141.
- Smyth JD, Dillman DA, Christian LM and Stern MJ (2006) Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly* 70(1): 66–77.
- Vannette DL and Krosnick JA (2014) Answering questions: A comparison of survey satisficing and mindlessness. In: Ie A, Ngunounen CT and Langer EJ (ed.) *The Wiley Blackwell Handbook of Mindfulness*. New York: John Wiley & Sons.
- Webb EJ, Campbell DT, Schwartz RD, Sechrest L and Grove JB (1981) *Nonreactive measures in the social sciences*. Boston, MA: Houghton Mifflin.
- Yeager DS and Krosnick JA (2010) The validity of self-reported nicotine product use in the 2001–2008 national health and nutrition examination survey. *Medical Care* 48(12): 1128–1132.