A Scent of Strategy: Response Error in a List Experiment on Anti-Immigrant Sentiment

Rinken, Sebastian; Pasadas-del-Amo, Sara; Trujillo-Carmona, Manuel

Empfohlene Zitierung / Suggested Citation:
Rinken, S., Pasadas-del-Amo, S., & Trujillo-Carmona, M. (2024). A Scent of Strategy: Response Error in a List Experiment on Anti-Immigrant Sentiment. *Methods, data, analyses : a journal for quantitative methods and survey methodology (mda)*, *18*(2), 249-262. https://doi.org/10.12758/mda.2024.05

methods, data, analyses | 2024, pp. 249-262 DOI: 10.12758/mda.2024.05

A Scent of Strategy: Response Error in a List Experiment on Anti-Immigrant Sentiment

Sebastian Rinken[1], Sara Pasadas-del-Amo[2] & Manuel Trujillo-Carmona[1]

[1] *Instituto de Estudios Sociales Avanzados (IESA), CSIC, Córdoba, Spain*

[2] *Universidad de Córdoba, Spain*

Abstract

This Research Note reports on a list experiment regarding anti-immigrant sentiment (n=1,965) that was fielded in Spain in 2020. Among participants with left-of-center ideology, the experiment originated a *negative* difference-in-means between treatment and control. Drawing on Zigerell's (2011) deflation hypothesis, we assess the possibility that leftist treatment group respondents may have altered their scores *by more than one* to distance themselves unmistakably from the sensitive item. We consider this possibility plausible in a context of intense polarization where immigration attitudes are closely associated with political ideology. This study's data speak to the results of recent meta-analyses that have revealed list-experiments to fail when applied to prejudiced attitudes and other highly sensitive issues – i.e., precisely the kind of issues with regard to which the technique ought to work best. We conclude that the possibility of strategic response error in specific respondent categories needs to be considered when staging and interpreting list experiments.

Keywords: item-count technique, social desirability bias, strategic response error, immigration attitudes, political polarization

The list experiment, or item-count technique (ICT), aims to obtain unbiased estimates of sensitive behaviors or attitudes. Respondents are divided randomly in treatment and control groups, administered identical lists except for the target item's addition as treatment, and asked *how many,* but not which, items apply to them. The sensitive item's prevalence is estimated by comparing both groups' differences-in-means (DiMs), and the extent of social desirability bias (SDB) assessed by contrast with an equally worded direct question (DQ) (Miller, 1984; Glynn, 2013). This paper dwells on a list experiment on anti-immigrant sentiment that obtained an apparently non-sensical *negative* difference-in-means for some respondents (but not others). Among participants with leftist ideology, the experiment's mean score was significantly *lower* when exposed to treatment (addition of "immigrants" as potentially antipathetic group) than when confronted only with an otherwise identical list of control items. The ensuing aggregate result echoes the findings of a recent meta-analysis that detects *reverse* ICT-DQ differences in studies of prejudiced attitudes (Blair et al., 2020); a second meta-analysis observes disappointing ICT results regarding highly sensitive items (Ehler et al., 2021). Our data offer a rare opportunity for exploring response patterns in specific participant categories, a line of research that might contribute to discerning why list experiments tend to fail precisely when applied to the kind of issues for which they ought to work best.

Background and Objectives

ICT has been employed to gauge the prevalence of ill-regarded behaviors and attitudes such as drug use, risky sex, vote buying, racism, or anti-Semitism, and well-regarded ones such as voting or charitable giving, among many others (Tourangeau & Yan, 2007; Holbrook & Krosnick, 2010; Krumpal, 2013; Blair et al., 2020). Four control items, one each of ample and scarce prevalence and two mutually exclusive ones, are recommended to prevent respondents from considering all items applicable (ceiling), or none (floor), situations that would compromise perceived anonymity (Kuklinski et al., 1997; Blair & Imai, 2012); sensitive controls should be avoided if possible (Droitcour et al., 1991; Ehler et al., 2021). ICT is generally rated as preferable to other unobtrusive survey pro-

Acknowledgements

We would like to thank all EASIE team members as well as IESA's Technical Unit for Applied Studies and Sara Varela (IMOP) for their varied contributions to the survey's successful implementation (cf. https://www.iesa.csic.es/en/proyectos/explaining-placid-attitudes-towards-immigrants-in-spain-easie/).

Direct correspondence to

Sebastian Rinken, Instituto de Estudios Sociales Avanzados (IESA), CSIC, Córdoba, Spain
E-mail: srinken@iesa.csic.es

cedures such as randomized response technique, which guarantees privacy by requesting a score for *either* the sensitive item *or* an unrelated one, for example – petitions that might confuse or even irritate some participants (Coutts & Jann, 2011; Hox & Lensvelt-Mulders, 2008; Rosenfeld et al., 2016; Wolter & Diekmann, 2021). Although list experiments are comparatively straightforward, a growing number of papers have voiced concerns about various kinds of non-strategic response error and ensuing instability (Tsuchiya & Hirai, 2010; Kiewiet de Jonge & Nickerson, 2014; Ahlquist, 2018; Gosen et al., 2019; Kramon & Weghorst, 2019; Jerke et al., 2019; Ehler et al., 2021; Kuhn & Vivyan, 2021; Riambau & Ostwald 2021; Jerke et al., 2022).

The list experiment's most notorious drawback is outsize variance (Miller, 1984; Blair et al., 2020; Ehler et al., 2021); Blair and colleagues (2020) estimate ICT to be 14 times (!) more variable than DQs. Hence, even for considerable differences vis-à-vis obtrusive measures, extremely large samples are required to clear customary significance thresholds. Since this problem is exacerbated in subgroups, little is known about the scope, or even direction, of ICT-DQ comparisons in specific respondent categories (Lax et al., 2016; Blair et al., 2020). A related hitch is relative opacity regarding covariates: vast standard errors arise when regressing ICT results on predictors (Corstange, 2009; Imai, 2011; Blair & Imai, 2012; Glynn, 2013).

Most list experiments obtain reduced bias as compared to obtrusive measurement. Recent meta-analyses conclude that ICT improves estimates of SDB-prone behaviors or mindsets by 8.5 (Ehler et al., 2021) to 10 percentage points (Blair et al., 2020) on average as compared to DQs. However, ICT's performance varies strongly across substantive domains (Blair et al., 2020). Startlingly, the technique has defied expectations with regard to highly sensitive items in general (Ehler et al., 2021) and prejudiced attitudes, in particular (Blair et al., 2020). Blair and colleagues (2020) even find ICT-based prejudice estimates to diverge from DQ-based ones in the *opposite* direction. How may such data be accounted for?

One possible explanation, the reverse polarity of social norms, has been documented in specific contexts, such as nativism in the US (Knoll, 2013), anti-immigrant sentiment in Japan (Igarashi & Nagayoshi, 2022), and vote-buying in Nigeria (Hatz *el al.*, 2023). However, reverse SDB seems implausible with regard to prejudiced attitudes and other highly sensitive items in general (since that proposition would presuppose the reverse polarity of social norms *tout court*), and it cannot possibly explain why treatment respondents mark *lower* scores than their control-group peers.

ICT's rationale relies on encouraging insincere norm violators to alter their score *by one* when faced with the sensitive item. Two crucial assumptions apply (Imai, 2011; Blair & Imai, 2012): sincere scores regarding the sensitive item ("no liars"), and indifference of control item scores to treatment ("no design effect"). Extant scholarship contemplates strategic response error almost exclusively with regard to the experiment's intended addressees (insincere norm violators), hence insisting on optimal anonymity safeguards (cf. ceiling/floor). However, the situation thus created may pose difficulties for respondents keen to distance themselves unequivocally from the sensitive item. This possibility –which seems especially plausible with regard to norm *adherers*– was first observed by Zigerell (2011, p. 553): to prevent any risk of being associated with the treatment item, some respondents may deflate their score "by *any number*", thereby originating *negative* differences between treatment- and control-group scores and distorting aggregate estimates of the sensitive item and related bias. Analogously, respondents keen to send an unmistakable signal of association with a socially desirable treatment item might inflate their scores *by more than one*. Such response behavior would constitute a "design effect" of sorts, yet one deriving from confrontation with the treatment item as such, rather than a flawed choice of controls. Apart from Zigerell's (2011) work on racism, deflation effects have been reported by just a handful of studies, all of which regard strongly polarizing issues such as marijuana use (García-Sánchez & Queirolo, 2020), violent extremism (Clemmow et al., 2020), or anti-immigrant sentiment (Rinken et al., 2021).

This study adds to the extant literature in three ways. First, we document a *negative* difference-in-means between treatment and control among respondents with leftist ideology – a rare opportunity to explore subgroup-level response behavior in a list experiment. Second, we argue that non-strategic error fails to explain why the *longer* list induces *lower* mean scores in this respondent category, but not others. This is important, given that such explanations are favored by the extant literature. Third, by building on Zigerell's (2011) work, we hypothesize various reasons for leftist respondents to deflate their experimental scores *by more than one* in the study context. Negative DiMs in participant subgroups entail an additional rationale, other than and potentially complementary to reverse SDB, for explaining *reverse* aggregate ICT-DQ differences (cf. Blair et al., 2020). Our data highlight the need for further research on the possibility of strategic response error in list experiments on prejudiced attitudes and other highly sensitive items.

Data and Method

A list experiment on anti-immigrant sentiment (AIS) was included in a web survey on native citizens' attitudes toward immigration and immigrants (see online appendix, Figures A1 through A3). Respondents were asked toward how many, among various social groups, they felt antipathy. "Immigrants" were added as treatment to four control items, two of which antagonist (labor unionists and multi-millionaires), one low-prevalence (compulsive gamblers) and one high-prevalence (drug dealers). Control-group respondents were subsequently asked heads-on about antipathy towards immigrants; random assignment to control or treatment ensures the comparability of both estimates. The term “antipathy” refers to the affective core of prejudiced attitudes (Allport, 1954) in a negatively charged way that seems prone to elicit desirability pressures. Hence, our baseline expectation was that the ICT estimate (DiM) would exceed the direct AIS gauge. Control items were chosen based on two pretests, one regarding the entire questionnaire (n=86) and a second one (n=220) focusing on ICT design (see section 1 of the online appendix for details). While the chosen list performed well, some pre-tested options originated negative DiMs – with hindsight, a bellwether of our study’s results.

The survey was administered in 2020 to an online sample of Spanish nationals born and resident in Spain (n=1,965). The sample was selected randomly from a probability-based online panel recruited via random digit dial surveys (see online appendix, Tables A1 and A2). Since we focus on comprehending the response patterns observed in this particular experiment rather than producing population estimates, we use unweighted data in this paper.

Randomization worked well: the covariate profiles of the experiment’s control and treatment arms are almost identical (see online appendix, Table A3). ICT non-response was negligible (1 and 2 persons respectively in treatment and control), and there are few cases at either tail of the item score distribution for both experimental groups, indicating that the experimental design avoided significant ceiling and floor effects (see online appendix, Figure A4). The test for design effects (Blair & Imai, 2012) was passed although a negative proportion is estimated for one respondent type (online appendix, Table A4). This does not prove the absence of design effects (Blair & Imai, 2012): rather, the test did not exclude the possibility of the negative value having arisen by chance. SDB was estimated with R-LIST as difference between a linear-model fit for the ICT and a logit-model fit for the DQ result (Blair & Imai, 2012). Covariates of the ICT-based AIS estimate were modeled by nonlinear least squares (NLS) and maximum likelihood (ML) regressions as implemented in R-LIST; covariates of manifest AIS were modeled as logit regression (Imai, 2011; Blair & Imai, 2012; Blair, Chou & Imai, 2018) (see online appendix for details).

Results

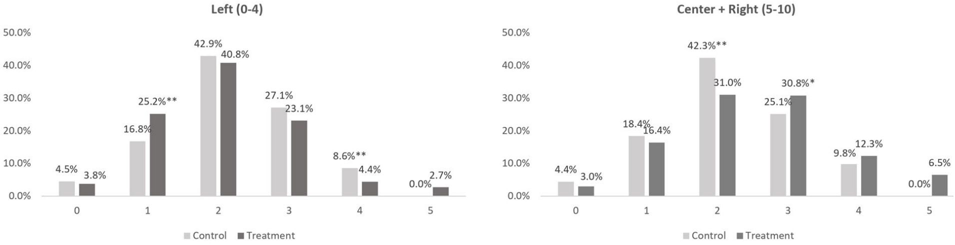
The experiment failed to generate the increased AIS estimate we had anticipated (Table 1). On aggregate, the treatment group’s mean score exceeds the control group’s mean, but the ensuing AIS estimate does not differ significantly from the DQ-based result even when lowering the customary 95% confidence interval (AIS range: 3% to 21.8%) to 90% (range: 4.5% to 20.3%). That said, the ICT-based estimate is actually 3.4 percentage points *lower* than the DQ-based one.

Table 1 Estimates of anti-immigrant sentiment (ICT vs. direct question) and SDB

	Control mean	Treatment mean	ICT estimate (DiM)	DQ estimate	SDB
	2.183	2.308	0.124	0.159	-0.034
S.E.	(0.031)	(0.036)	(0.048)	(0.012)	(0.049)
N	974	988		973	

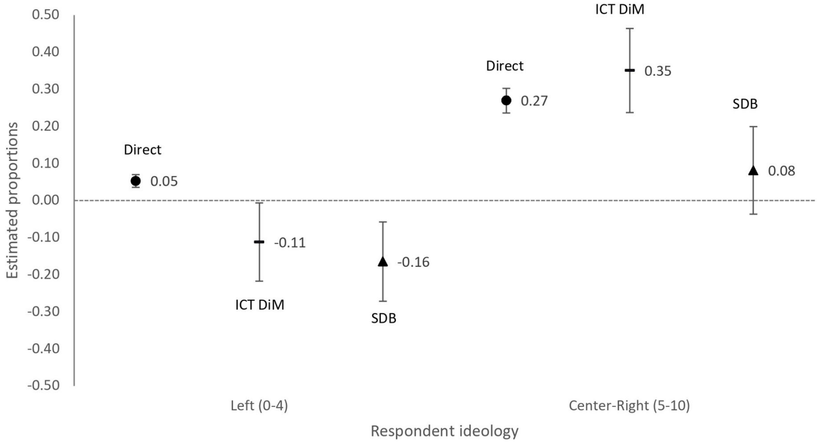
Source: EASIE survey. Abbreviations: ICT=Item-count technique; DiM=difference-in-means between control and treatment; DQ=direct question; SDB=social desirability bias (difference between ICT and DQ-based estimates).

Closer inspection reveals that the experiment generated different response patterns in distinct participant categories. This situation, discernible for several sociodemographic variables including educational attainment and age group, is observed most clearly with regard to political ideology. Treatment participants with centrist or right-of-center ideology mark less 2s and increasing proportions of higher scores (especially 3s) than their control group peers. However, among leftist treatment respondents, the share of 1s increases significantly by comparison to the control group, whereas the proportions of higher scores (especially 4s) decrease (Figure 1). Consequently, among respondents with centrist or right-of-center ideology, our ICT-based estimate of anti-immigrant sentiment exceeds the direct gauge by about 8 percentage points, a non-significant difference. In sharp contrast, an AIS estimate of *minus* 11%, as opposed to 5% in DQ, is obtained for respondents with left-of-center ideology (significant for 90% confidence interval) (Figure 2).



Source: EASIE survey. (Left Total=963, Control= 487 Treatment= 476; Center-right Total= 984, Control= 478, Treatment= 506). Categories of political ideology were derived from self-ratings on a 0-10 scale where ‘0’ means ‘completely leftist’ and ‘10’ means ‘completely rightist’. * p < 0.05; ** p< 0.01.

Figure 1 Item scores in list experiment on anti-immigrant sentiment (unweighted), by respondent ideology



Source: EASIE survey. (Left Total=963, Control= 487 Treatment= 476, DQ=487; Center-right Total= 984, Control= 478, Treatment= 506; DQ=477). Categories of political ideology were derived from self-ratings on a 0-10 scale where ‘0’ means ‘completely leftist’ and ‘10’ means ‘completely rightist’. Bars represent 90% confidence intervals.

Figure 2 Estimates of anti-immigrant sentiment (DQ vs. ICT-DiM) and social desirability bias, by political ideology

Political ideology is a consistent predictor of immigration attitudes (Ceobanu & Escandell, 2010; Hainmueller & Hopkins, 2014; Dražanová, 2022): leftist ideology is generally associated with more benevolent views, and rightist ideology with more restrictive or intolerant ones. In our study, political ideology is associated, net of sociodemographic controls, to both AIS gauges (see online appendix, Table A5). Our study is not powered to assess DiM estimates for each point of the ideological self-rating scale, but those data (see Figure A6 in the online appendix) clearly support the creation of the two groupings (0-4 vs. 5-10) considered here.

Discussion

This study is hampered by ICT’s notorious weakness of large variance. The *negative* aggregate difference vis-à-vis direct measurement and the *positive* difference among respondents with centrist or right-of-center ideology both fail to clear any meaningful significance threshold, and the 11-points *negative* difference-in-means between treatment and control among participants with left-of-center ideology is significant only for a 90% confidence interval. This situation might tempt some analysts to dismiss the data as spurious. However, it seems worth noting that our study’s aggregate result echoes the *opposite* margin vis-à-vis DQs detected by a recent meta-analysis in list experiments on prejudiced attitudes (Blair et al., 2020); another meta-analysis reveals disappointing ICT results when applied to highly sensitive items (Ehler et al., 2021). While desirability pressures might in some cases be trivial enough for obtrusive measurement to capture such items reasonably well, it seems precipitated to extend that hypothesis to prejudiced attitudes in general (Blair et al., 2020: 1310), and it seems implausible to attribute the inverse relation between item sensitivity and ICT effects (Ehler et al., 2021) to reverse SDB. Negative DiMs in sample subgroups caution against such interpretations. From this perspective, our data offer a welcome opportunity for exploring why ICT seems prone to fail precisely when it ought to work best. Given these circumstances, we consider a suboptimal significance level justified here. Hence, in the following, we will dwell on possible reasons for leftist treatment respondents to mark *lower* mean scores than their control group peers.

Most extant scholarship attributes counter-intuitive or inconclusive ICT data to various kinds of non-strategic response error, such as comprehensibility issues (Kramon & Weghorst, 2019; Jerke et al., 2019), unequal length of lists (Tsuchiya & Hirai, 2010), perceived weirdness (Kuha & Jackson, 2014), or confounding control items (Ehler et al., 2021). We find these explanations unconvincing with regard to our data. Given the negligible incidence of non-response, we see no reason to suspect that the experiment posed excessive cognitive difficulties. Actually, negative DiMs are observed across education levels among leftist respondents (however, large variance keeps these results from attaining statistical significance). If a higher number of items, as such, were to distort results, we see no reason why this should apply only to participants with left-of-center ideology. Similarly, if erratic responses were occasioned by the potentially disconcerting nature of the experimental task (“just how many…”), they should occur regardless of participants’ ideological profiles. In both ideological groupings, response times of treatment participants increased by almost identical margins (4.9 and 4.7 seconds, respectively) as compared to controls (see online appendix, Table A6); given the need to consider a higher number of items, such an increase should be expected. However, respondent ideology might come into play with regard to control items. To prevent ceiling and floor effects, lists are required to contain two mutually exclusive items (Kuklinski et al., 1997; Blair & Imai, 2012). When inquiring about antipathy toward a varied assortment of social groups, it seems inevitable that one such might be perceived as sensitive by some respondents; specifically, in our study, some leftist respondents might have been reluctant to admit antipathy toward labor unionists. If so, though, both experimental groups should be similarly affected by such reluctance. Therefore, we do not see how the treatment arm’s *lower* mean could derive from desirability pressures regarding labor unionists.

Bearing in mind that the experiment is exactly the same for all participants, *except for inclusion of an additional item as treatment*, confrontation with this item offers the most straightforward explanation for any differential response pattern vis-à-vis control. Indeed, the expectation that list experiments ought to originate improved prevalence estimates of sensitive behaviors and attitudes is predicated on this premise: norm violators are supposed to alter their score by one (by comparison to analogous control group participants) when faced with the treatment item, whereas all other treatment respondents are supposed to be unaffected by the sensitive item. However, treatment participants who fervently adhere to the norm might react in unanticipated ways, as might stubbornly insincere norm violators. The possibility that the experimental situation might originate strategic response error has played a subdued role in the scholarly debate thus far. In an apparent nod to Zigerell's (2011) work on racism, Blair et al. (2020: 1310) acknowledge passingly "that the list experiment (might) not provide the cover it is designed to provide in this context", yet do not elaborate any further.

The experimental situation's opacity ("*just how many*") is meant to encourage insincere norm violators to lower their guard. Zigerell (2011) argued that this very opacity may prove challenging for respondents aiming to send a clear signal of dissociation from a negatively charged item. He hypothesized that such respondents may alter their score *by more than one*, thereby originating a negative DiM by comparison to their control-group peers (an analogous logic of "overacting" may apply to positively charged treatment items). Such deflation effects presuppose very intense desirability pressures, as was the case with Zigerell's data on racism in the U.S. Because unwelcoming attitudes toward immigrants are prone to be interpreted as telltale of racist or xenophobic views (Esses et al., 1998; Wilkes et al., 2008), the possibility of similarly intense desirability dynamics seems worth considering here. With regard to AIS in Spain, deflation effects have been documented among self-declared xenophiles (Rinken et al., 2021), who are by definition keen to distance themselves from anti-immigrant prejudice. Since attitudes toward immigration and immigrants correlate strongly with political ideology (in our dataset, correlation coefficients exceed 0.38 for various ATII gauges), it seems fair to assume that leftist respondents and self-declared xenophiles react similarly to a list experiment on AIS. However, in our study, negative DiMs are statistically significant for leftist respondents but not for xenophile ones; this situation suggests some additional factor driving leftist participants' response behavior.

The empirical context of our study entails discernible incentives, apart from and beyond xenophile attitudes, for leftist respondents to seek clear dissociation from AIS. For the first time since the Franco dictatorship's demise, a radical-right party featuring anti-immigrant rhetoric had recently scored significant electoral gains across Spain (Ferreira, 2019; Mendes & Dennison, 2021; Turnbull-

Dugarte et al., 2020). Consequently, immigration-related issues became supercharged ever more intensely by broader questions of ideological allegiance. This context is reflected by intensifying polarization of survey data on immigration attitudes (González Enríquez & Rinken, 2021): in direct measurement, right-wing respondents manifest increasingly unfavorable views, whereas leftwingers state increasingly favorable positions. Such data might reflect genuine trends (souring or improved attitudes, respectively), but enhanced desirability pressures might play a role too. To participants with right-wing ideology who pay lip-service to anti-immigrant rhetoric in DQ, the list experiment offers the coverage needed for revealing their true feelings.

In contrast, leftists are in a bind. In our survey's context of intense ideological polarization, it seems plausible to assume that the experimental situation might be experienced as inconveniently opaque by some leftist participants. Whatever their mindset regarding immigration and immigrants, this context makes it tempting for leftist treatment-group respondents to distance themselves unmistakably from a sensitive item that is routinely tagged, in their ideological ecosystem, as deplorable epitome of right-wing extremism. The ensuing scores do not reveal true feelings: leftist treatment respondents might opt to deflate their scores either because of particularly strong xenophile convictions, or else due to an intense wish to appear to be sharing such convictions. Our data offer no insight about the relative importance of either, but raw scores do indicate a constraint (cf. Figure 1): an overwhelming majority mark scores higher than zero. Thus, the urge to unmistakably flag anti-racist convictions does not propel leftist treatment respondents to induce any doubt about their disdain for drug-dealers. Also, it seems worth noting that the data indicate a minimum level of deflation behavior, rather than measuring its exact extent: a negative difference-in-*means* is observed net of the *increased* scores that some leftist participants may have marked when faced with the treatment item.

Who might such advertisements of norm compliance be directed to? Response behavior in survey settings is meaningful only with regard to an (imaginary or tangible) audience. Most SDB studies have considered external audiences, such as interviewers or bystanders; however, recent research retrieves interest in the self as ever-present and potentially decisive audience (Blair et al., 2020; Brenner, 2020). In our panel-based data, survey administrators cannot be discarded as salient social referent (Coutts & Jann, 2011) – be it to safeguard one's xenophile credentials, or else to counterfactually exhibit politically correct attitudes. Yet, the experimental situation may also induce respondents to "edit their report for their own benefit; that is, for their own view of themselves" (Brenner, 2020: 49). In a context of strong polarization regarding immigration-related issues, it seems plausible that ideological self-identifications may claim center stage. Thus, leftist treatment respondents may alter their list scores *by more than one* either to burnish a genuine self-image of benevolence towards immigrants, or else to dispel the dissonant chord (Festinger, 1957) struck by the sensitive item with regard to their overall ideology. Cross-tabulation of both parameters (relation with the pro-immigrant norm, on one hand, and projected audience, on the other) originates a tentative taxonomy of deflation motives that might benefit future attempts at refining their conceptualization (see online appendix, Table A7).

Conclusion

This exploratory study aims to stimulate further research on the methodological properties of list experiments. Apart from heeding the recommendation to field future list experiments with extremely large samples, survey methodologists and practitioners interested in highly sensitive issues should consider two related possibilities: (a) inconclusive or counter-intuitive aggregate data might stem from divergent response patterns in participant subgroups, and (b) strategic response error might contribute to their explanation.

Funding

This work was supported by grant CSO2017-87364-R, funded by the Spanish Ministry of Science and Innovation (MCIN/ AEI /10.13039/501100011033/) and the European Regional Development Fund (ERDF “A way of making Europe”).

Ethics approval

The dataset used in this study was generated by a survey approved by the Spanish Research Council’ Ethics Committee (reference nº 127/2020).

Data availability statement

The dataset is available at CSIC’s institutional open-access repository (cf. Rinken et al., 2023).

References

Ahlquist, J. S. (2017). List experiment design, non-strategic respondent error, and item count technique estimators. *Political Analysis, 26,* 34–53. https://doi.org/10.1017/pan.2017.31

Allport, G. (1954). *The nature of prejudice*. Reading: Addison-Wesley.

Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis, 20*, 47–77. https://doi.org/10.1093/pan/mpr048

Blair, G., Chou, W., & Imai, K. (2019). List experiments with measurement error. *Political Analysis, 27*, 455-480. https://doi.org/10.1017/pan.2018.56

Blair, G., Coppock, A., & Moor, M. (2020). When to worry about sensitivity bias: A social reference theory and evidence from 30 years of list experiments. *American Political Science Review. 114*, 1297–1315 https://doi.org/10.1017/s0003055420000374

Brenner, P. S. (2020). Advancing theories of socially desirable responding: How identity processes influence answers to 'sensitive questions'. In P. S. Brenner (Ed.), *Understanding survey methodology: sociological theory and applications* (pp. 45–65). Cham: Springer. https://doi.org/10.1007/978-3-030-47256-6_3

Ceobanu, A. M., & Escandell, X. (2010). Comparative analyses of public attitudes toward immigrants and immigration using multinational survey data: A review of theories and research. *Annual Review of Sociology, 36,* 309–28. https://doi.org/10.1146/annurev.soc.012809.102651

Clemmow, C., Schumann, S., Salman, N. L., & Gill, P. (2020). The base rate study: Developing base rates for risk factors and indicators for engagement in violent extremism. *Journal of Forensic Sciences, 65*, 865–81. https://doi.org/10.1111/1556-4029.14282

Corstange, D. (2009). Sensitive questions, truthful answers? Modeling the list experiment with LISTIT. *Political Analysis, 17,* 45–63. https://doi.org/10.1093/pan/mpn013

Coutts, E., & Jann, B. (2011). Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). *Sociological Methods & Research, 40*(1), 169–93. https://doi.org/10.1177/0049124110390768

Dražanová, L. (2022). Sometimes it is the little things: A meta-analysis of individual and contextual determinants of attitudes toward immigration (2009–2019). *International Journal of Intercultural Relations, 87,* 85–97. https://doi.org/10.1016/j.ijintrel.2022.01.008

Droitcour, J., Caspar, R. A., Hubbard, M. L., Parsley, T. L., Visscher, W., & Ezzati, T. M. (1991). The item count technique as a method of indirect questioning: A review of its development and a case study application. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Thiowetz, & S. Sudman (eds.) *Measurement errors in surveys* (pp. 185–210). John Wiley & Sons. https://doi.org/10.1002/9781118150382.ch11

Ehler, I., Wolter. F., & Junkermann, J. (2021). Sensitive questions in surveys. A comprehensive meta-analysis of experimental survey studies on the performance of the item count technique. *Public Opinion Quarterly, 85*(1), 6–27. https://doi.org/10.1093/poq/nfab002

Esses, V., Jackson, L., & Armstrong, T. (1998). Intergroup competition and attitudes toward immigrants and immigration: An instrumental model of group conflict. *Journal of Social Issues, 54*(4), 699-724. https://doi.org/10.1111/j.1540-4560.1998.tb01244.x

Ferreira, C. (2019). Vox as representative of the radical right in Spain: A study of its ideology. *Revista Española de Ciencia Política,* (51), 73–98. https://doi.org/10.21308/recp.51.03

Festinger, L. (1957). *A theory of cognitive dissonance,* Stanford, CA: Stanford University Press. https://doi.org/10.1515/9781503620766

García-Sánchez, M., & Queirolo, R. (2020). A tale of two countries: The effectiveness of list experiments to measure drug consumption in opposite contexts. *International Journal of Public Opinion Research, 33*(2), 255–72. https://doi.org/10.1093/ijpor/edaa031

Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly, 77*(S1), 159–72. https://doi.org/10.1093/poq/nfs070

González Enríquez, C., & Rinken, S. (2021). Spanish public opinion on immigration and the effect of VOX. ARI 46/2021. Madrid: Real Instituto Elcano.

Gosen, S., Schmidt, P., Thörner, S., & Leibold, J. (2019). Is the list experiment doing its job? In J. Mayerl, T. Krause, A. Wahl, & M. Wuketich (eds.), *Einstellungen und Verhalten in der empirischen Sozialforschung: Analytische Konzepte, Anwendungen und Analyseverfahren* (pp. 179–205). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-16348-8_8

Hainmueller, J., & Hopkins, D. J. (2014). Public attitudes toward immigration. *Annual Review of Political Science, 17*(1), 225–49. https://doi.org/10.1146/annurev-polisci-102512-194818

Hatz, S., Fjelde, H. & Randahl, D. (2023). Could vote buying be socially desirable? Exploratory analyses of a 'failed' list experiment. *Quality & Quantity*. https://doi.org/10.1007/s11135-023-01740-6

Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports tests using the item count technique. *Public Opinion Quarterly, 74*(1), 37–67. https://doi.org/10.1093/poq/nfp065

Hox, J. & Lensvelt-Mulders, G. (2008). Randomized response. In Lavrakas, P. J. *Encyclopedia of survey research methods*. Sage publications. https://dx.doi.org/10.4135/9781412963947

Igarashi, A., & Nagayoshi, K. (2022). Norms to be prejudiced: List experiments on attitudes towards immigrants in Japan. *Social Science Research*, 102, 102647. https://doi.org/10.1016/j.ssresearch.2021.102647

Imai, K. (2011). Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association, 106*(494), 407–16. https://doi.org/10.1198/jasa.2011.ap10415

Jerke, J., Johann, D., Rauhut, H., & Thomas, K. (2019). Too sophisticated even for highly educated survey respondents? A qualitative assessment of indirect question formats for sensitive questions. *Survey Research Methods, 13*, 319–51. https://doi.org/10.18148/srm/2019.v13i3.7453

Jerke, J., Johann, D., Rauhut, H., Thomas, K., & Velicu, A. (2022). Handle with care: implementation of the list experiment and crosswise model in a large-scale survey on academic misconduct. *Field Notes, 34*(1), 69–81. https://doi.org/10.1177/1525822x20985629

Kiewiet de Jonge, C. P., & Nickerson, D. W. (2014). Artificial inflation or deflation? Assessing the item count technique in comparative surveys. *Political Behavior, 36*(3), 659–82. https://doi.org/10.1007/s11109-013-9249-x

Knoll, B. R. (2013). Implicit nativist attitudes, social desirability, and immigration policy preferences. *International Migration Review*, 47(1), 132-165. http://dx.doi.org/10.1016/j.ssresearch.2013.07.012

Kramon, E., & Weghorst, K. (2019). (Mis)Measuring sensitive attitudes with the list experiment. *Public Opinion Quarterly, 83*(S1), 236–63. https://doi.org/10.1093/poq/nfz009

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity, 47,* 2025–47. https://doi.org/10.1007/s11135-011-9640-9

Kuha, J., & Jackson, J. (2014). The item count method for sensitive survey questions: Modelling criminal behaviour. *Journal of the Royal Statistical Society, Series C (Applied Statistics), 63(2),* 321-341. https://doi.org/10.2139/ssrn.2119238

Kuhn, P. M., & Vivyan. N. (2021). The misreporting trade-off between list experiments and direct questions in practice: Partition validation evidence from two countries. *Political Analysis,* published online April 16, 2021. https://doi.org/10.1017/pan.2021.10

Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial attitudes and the 'New South'. *The Journal of Politics, 59*(2), 323–49. https://doi.org/10.2307/2998167

Lax, J. R., Phillips, J. H., & Stollwerk, A. F. (2016). Are survey respondents lying about their support for same-sex marriage? Lessons from a list experiment. *Public Opinion Quarterly, 80*(2), 510–33. https://doi.org/10.1093/poq/nfv056

Mendes, M. S., & Dennison, J. (2021). Explaining the emergence of the radical right in Spain and Portugal: Salience, stigma and supply. *West European Politics, 44*(4), 752–75. https://doi.org/10.1080/01402382.2020.1777504

Miller, J. D. (1984). A new survey technique for studying deviant behavior. PhD Thesis, George Washington University.

Riambau, G., & Ostwald, K. (2021). Placebo statements in list experiments: Evidence from a face-to-face survey in Singapore. *Political Science Research and Methods, 9*(1), 172–79. https://doi.org/10.1017/psrm.2020.18

Rinken, S., Pasadas-del-Amo, S., Rueda, M., & Cobo, B. (2021). No magic bullet: Estimating anti-immigrant sentiment and social desirability bias with the item-count technique. *Quality & Quantity, 55*, 2139–59. https://doi.org/10.1007/s11135-021-01098-7

Rinken, S., Buraschi, D., Domínguez Álvarez, J. A., Godenau, D., González Enríquez, C., Lafuente, R., ... Varela, S. (2023). Survey on attitudes toward immigration and immigrants in Spain (EASIE survey) [Data set]. DIGITAL.CSIC. https://doi.org/10.20350/DIGITALCSIC/15586

Rosenfeld, B., Imai, K., & Shapiro, J.N. (2016). An empirical validation study of popular survey methodologies for sensitive questions. *American Journal of Political Science, 60*(3), 783–802. https://doi.org/10.1111/ajps.12205

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859-883. https://doi.org/10.1037/0033-2909.133.5.859

Tsuchiya, T., & Hirai, Y. (2010). Elaborate item count questioning: Why do people underreport in item count responses? *Survey Research Methods, 4*(3), 139-149. https://doi.org/10.18148/srm/2010.v4i3.4620

Turnbull-Dugarte, S. J., Rama, J., & Santana, A. (2020). The Baskerville's Dog suddenly started barking: Voting for VOX in the 2019 Spanish general elections. *Political Research Exchange* 2(1), 1781543. https://doi.org/10.1080/2474736x.2020.1781543

Wilkes, R., Guppy, N., & Farris, L. (2008). No thanks, we're full: Individual characteristics, national context, and changing attitudes toward immigration. *International Migration Review, 42*(2), 302-329. https://doi.org/10.1111/j.1747-7379.2008.00126.x

Wolter, F. and Diekmann, A. (2021). False positives and the 'more-is-better' assumption in sensitive question research. *Public Opinion Quarterly, 85*(3), 836–63. https://doi.org/10.1093/poq/nfab043

Zigerell, L. J. (2011). You wouldn't like me when I'm angry: List experiment misreporting. *Social Science Quarterly, 92*(2), 552–62. https://doi.org/10.1111/j.1540-6237.2011.00770.x