International Journal of Public Opinion Research

doi:10.1093/ijpor/edw002

RESEARCH NOTE

Estimating Sensitive Behavior: The ICT and High-Incidence Electoral Behavior

Kathrin Thomas[1,2], David Johann[1,3], Sylvia Kritzinger[1], Carolina Plescia[1], and Eva Zeglovits[1,4]

[1]Department of Methods in the Social Sciences, University of Vienna, Austria
[2]Centre for Comparative Social Surveys, City University London, United Kingdom
[3]German Centre for Research on Higher Education and Science Studies (DZHW)
[4]IFES, Austria

When respondents intentionally present themselves in socially desirable ways, survey researchers speak of *social desirability bias* (SDB) that is introduced to the data. SDB is a challenge to survey research, as it can lead to inaccurate measurements and incorrect conclusions about citizens' behavior (Phillips & Clancy, 1972). The *Item Count Technique* (ICT) is a frequently used method in survey research to reduce such bias by indirectly, anonymously, and confidentially asking respondents about the sensitive behavior in question.

Extensive social science research on the functioning of the ICT has compared its estimates with the measures obtained by traditional question formats that directly ask about citizens' behavior (Droitcour et al., 1991; Wolter & Laier, 2014). But, while practical applications and statistical refinements are already discussed (Blair & Imai, 2012; Corstange, 2009; Glynn, 2013; Imai, 2011; Tsuchiya, 2005), general questions with regard to the performance of the ICT remain unanswered: Approximately 80% of the existing studies have explored the ICT with regard to underreporting of socially *un*desirable attitudes when low-incidence behavior is concerned (i.e., behavior that rarely occurs, such as shoplifting, drug use, etc.). These studies obtained better (more accurate) ICT results compared with traditional direct question formats

All correspondence concerning this article should be addressed to Sylvia Kritzinger University of Vienna, Department of Methods in the Social Sciences, Rathausstraβe 19, 1010 Vienna - Austria. E-mail: sylvia.kritzinger@univie.ac.at

(Wolter & Laier, 2014). However, recent work has questioned the performance of the ICT in tackling SDB in high-incidence behavior (i.e., fairly common behavior that frequently occurs, such as voting; de Jonge & Nickerson, 2014). It is still unclear *whether* the ICT corrects for SDB when high-incidence behavior is concerned.

This article contributes to the recent strand of research by examining whether the ICT provides better estimates for socially desirable and for socially undesirable high-incidence behavior in comparison with direct question formats. More precisely, we test the performance of the ICT on high-incidence behavior in the field of elections and voting. We focus on Austria, where electoral turnout is high and radical right voting frequently occurs. While electoral turnout classifies as socially desirable behavior (Holbrook & Krosnick, 2010; Karp & Brockington, 2005; Zeglovits & Kritzinger, 2014), radical right voting describes socially undesirable behavior (Aichholzer, Kritzinger, Wagner, & Zeglovits, 2014).

The ICT's Power in Reducing SDB in High-Incidence Behavior

The ICT aims at reducing social desirability pressures by indirectly and confidentially asking respondents about sensitive behavior.[1] It is based on split ballots, where one half of the respondents are randomly assigned to a treatment group and the other half to a control group. Both groups are asked the same question, but while the treatment group is presented with a list of unobtrusive items *plus* the sensitive one, the sensitive behavior is removed from the control group's list. The technique ensures confidentiality and anonymity by asking respondents *how many* of the activities they have previously engaged in, so that they do not have to reveal which particular behavior they have performed. The sensitive item is covered up by the other unobtrusive behaviors. The ICT should therefore encourage respondents to more honest self-reports (Holbrook & Krosnick, 2010). The proportion of people engaging in the sensitive behavior is calculated by taking the mean difference of both lists and multiplying it by 100 (Kuklinski, Cobb, & Gilens, 1997).

The ICT has frequently been used in survey research across disciplines. For example, it has been applied in social psychology to study theft from employers, shoplifting, and anti-gay hate crime (Dalton, Wimbush, & Daily, 1994; Rayburn, Earleywine, & Davison, 2003; Tsuchiya, Hirai, & Ono, 2007), in health care research to analyze risky sexual behavior (LaBrie & Earleywine, 2000), and in political science to examine affirmative action for African Americans, feelings toward candidates' religion, race, or gender, vote buying, and voter turnout (Biemer & Brown, 2005; Çarkoğlu & Aytaç, 2015; Diaz-Cayeros et al., 2011; Flavin & Michael Keane, 2009; Gilens, Sniderman, & Kuklinski, 1998; Gonzalez-Ocantos, de Jonge, Meléndez, Osorio, & Nickerson, 2012; Heerwig & McCabe, 2009; Holbrook & Krosnick, 2010; Kane, Craig, & Wald, 2004; Redlawsk, Tolbert, & Franko, 2010; Streb, Burrell, Frederick, & Genovese, 2008).

[1]Admittedly, adjunct approaches as well as extensions to the traditional ICT have been developed. In particular, we would like to acknowledge different kinds of randomized (Coutts & Jann, 2011; Lensvelt-Mulders, Hox, van der Heijden, & Maas, 2005) and nonrandomized (Nepusz, Petróczi, Naughton, Epton, & Norman, 2014; Petróczi et al., 2011) response techniques. We focus on the traditional ICT as we would like to compare our findings with previous research relying on this technique.

Most studies attempt to tackle SDB in underreporting and compare the ICT's estimates with the results of the respondents' direct self-reports, assuming that the ICT should provide higher estimates ("more-is-better" assumption; Lensvelt-Mulders, Hox, van der Heijden, & Maas, 2005; Umesh & Peterson, 1991; Wolter & Laier, 2014). By providing an anonymous and indirect way of reporting socially undesirable activities, this assumption makes sense, as respondents should be encouraged to report their true behavior. Based on this assumption, we may conclude that the reverse relationship applies to overreporting. Citizens overreport behavior to present themselves as good citizens in front of others, even though they may not have participated in an activity. For example, it is socially accepted and admirable to contribute to a good cause (e.g., donations) or to comply with social norms and duties (e.g., electoral turnout). However, often the actual number of people who take part in these activities is lower and the true prevalence is known from official statistics. We may thus assume that in cases where overreporting applies, any result that is lower than the directly reported behavior should be more accurate ("less-is-better" assumption).

Previous evidence on the performance of the ICT when tackling underreporting is encouraging, given that the technique provides better (= higher) estimates of sensitive behavior compared with traditional question formats (de Jonge & Nickerson, 2014; Tsuchiya et al., 2007; Umesh & Peterson, 1991; Wolter & Laier, 2014). However, criticism about the ICT's empirical properties has been voiced across disciplines: scholars have paid attention to different conditions under which surveys were administered and which may explain why the ICT performs better in some surveys, but not in others. For example, the ICT may be sensitive to mode effects (Holbrook & Krosnick, 2010), sample sizes (Ahart & Sackett, 2004; Coutts & Jann, 2011; Dalton et al., 1994; Droitcour et al., 1991; Martinez & Craig, 2010), question formats (Corstange, 2009; also Glynn, 2013; Imai, 2011; Tsuchiya, 2005), and the selection of unobtrusive items (Droitcour et al., 1991; Kuklinski et al., 1997, Tsuchiya et al., 2007).[2] In addition, the robustness of the techniques that help validating the ICT's estimates also seems to be important.[3]

Limited attention has also been paid to the different kinds of sensitive behavior, particularly whether the assumption made by the ICT design applies to high- and low-incidence behavior in the same way. Indeed, the ICT performs well when it is used to reduce SDB of socially undesirable and low-incidence behavior and when validation is limited to comparing the ICT's estimates with direct self-reports (de Jonge & Nickerson, 2014). However, few studies have examined the value of the ICT

[2]For an overview of the broader debate, see also Andrew Gelman's blog "Thinking of doing a list experiment? Here's a list of reasons why you should think again." Retrieved May 29, 2015 from http://andrewgelman.com/2014/04/23/thinking-list-experiment-heres-list-reasons-think/.

[3]Research on underreporting often provides a weak validation test, as it solely relies on comparing the ICT's estimates with directly self-reported behavior. Studies on overreporting frequently have the benefit of the known prevalence of a reported behavior: Comparing the ICT's results against this known true prevalence, also referred to as a "strong validation study," allows to immediately adjudicate the validity of the ICT's results given that all other sources of error are limited to a minimum (Moshagen, Hilbig, Erdfelder, & Moritz, 2013). However, while such a comparison may seem to be a stronger approach in terms of validating the ICT's power to tackle SDB at first sight, other sources of error may apply, such as noncoverage, nonresponse, noncompliance, or other kinds of sampling errors (Nepusz et al., 2014; Holbrook & Krosnick, 2010; Traugott & Katosh, 1979).

in tackling SDB of socially desirable or undesirable *high*-incidence behavior, that is, activities that occur more frequently but in which not all citizens take part. When more frequent behavior is concerned, it is easier to compare the ICT's estimates with direct self-reports and real-world measures. Arguably, the assumptions with respect to the ICT's performance in reducing SDB on socially undesirable low-incidence behavior may not be applicable to socially (un-)desirable high-incidence behavior (de Jonge & Nickerson, 2014). For instance, prior work suggests that different mechanisms may be at work when underreporting of *socially undesirable* and overreporting of *socially desirable* behavior is concerned (Holbrook & Krosnick, 2010).

Citizens' electoral behavior is one good example for high-incidence behavior: Most people frequently participate in elections for different reasons, that is, duty, rational choice, or opinion expression (=socially desirable behavior; Blais, 2010). In addition, an increasing number of people across countries (e.g., Austria, France, the Netherlands) vote for parties located on the far right ideological spectrum (=socially undesirable behavior). Prior studies suggest that especially electoral turnout has a high probability of overreporting owing to SDB (Karp & Brockington, 2005). One explanation for the psychological mechanisms behind this is that respondents overreporting electoral behavior are similar to citizens who are already predisposed to vote (e.g., well-educated, dutiful, efficacious). Thus, these people are more likely to be aware of socially acceptable behavior and try to comply with social norms (Bernstein, Chadha, & Montjoy, 2001; Karp & Brockington, 2005; Silver, Anderson, & Abramson, 1986).

Indeed, the ICT has previously been applied in this field, but the evidence on its performance is mixed. Holbrook and Krosnick (2010) have been able to successfully reduce SDB in overreporting of turnout using telephone surveys, but have been unable to achieve similar results in online questionnaires. Similarly, Comşa and Postelnicu's (2013) study fails to obtain conclusive results. While it shows that the ICT performs better in comparison with the direct self-reports, the estimate is still far off the actual election results. In other words, "it would be improper to conclude that this [ICT] estimate is necessarily accurate" (Umesh & Peterson, 1991, p. 130). In situations where turnout validation, as is frequently practiced in the United States, is not an option, the ICT might yet be useful to capture electoral participation.[4]

Spinning this argument further, voting for radical right parties is a good example of socially undesirable high-incidence behavior, which is prone to underreporting (Aichholzer et al., 2014).[5] A similar underlying psychological mechanism is conceivable: Radical right voting is most prominent among young, less-educated, and male citizens (Johann, Thomas, Faas, & Fietkau, 2015; Werts, Scheepers, & Lubbers, 2013), but their

[4]Other research based on an adjunct technique to the ICT has been more successful in this field: A study by Moshagen, Musch, and Erdfelder (2012) was able to more accurately predict nonvoting in Germany (and thus also electoral participation). We do not follow up on Moshagen et al.'s (2012) findings in more detail, as their study does not rely on the ICT as such, but on a variant of the randomized response technique, the so-called stochastic lie detector.

[5]Electoral researchers are often confronted with low case numbers of radical right voters in surveys. However, elections statistics often indicate a higher proportion of radical right voters. To measure right-wing voting, scholars often rely on the so-called propensity to vote questions (PTVs) ("How likely is it that you will ever vote for [party]?," including radical right parties). While the PTVs provide higher case numbers and are frequently used to capture respondents' real voting behavior, their empirical properties have not been sufficiently tested yet. A first attempt of a cross-validation has been made by Johann, Thomas, Fass, and Fietkau (2015).

counterparts may still consider a radical right vote as unacceptable by social norms. As a consequence, these people may be more inclined to report that they have voted for another party to comply with the social norm that far right voting is unacceptable, even though they may have cast a ballot for a radical right party. Extreme statements with regard to immigration and integration, especially by Austria's radical right party, the Freedom Party of Austria (FPÖ), underline that radical right voting has a negative connotation in Austria and is thus prone to SDB.

We test the premise of high-incidence behavior in the two fields outlined above by examining the ICT's ability in reducing overreporting of electoral turnout (*socially desirable high*-incidence behavior) and underreporting of voting for radical right parties (*socially undesirable high*-incidence behavior). Our research therefore contributes to a better understanding of the conditions under which the ICT works in tackling SDB. We ask whether (1) the ICT is able to reduce SDB in high-incidence behavior (e.g., electoral behavior) following the same assumptions that apply to low-incidence behavior (e.g., counter intuitive behavior). Further, we investigate whether (2) the ICT performs equally well in capturing SDB for socially desirable and undesirable behaviors.

Three Survey Experiments

We implement the ICT in three different survey experiments. *Studies A* and *B* examine socially desirable behavior by addressing overreporting of electoral turnout. *Study C* addresses socially undesirable behavior in capturing underreporting of radical right voting. As all studies tap into electoral behavior, we are able to rely on the same unobtrusive items to hide our sensitive behaviors. Previous studies have provided us with two general guidelines on the selection of the unobtrusive items: To adequately hide the sensitive behavior among the unobtrusive activities, (1) the items should share a common theme (Dalton et al., 1994) and (2) they should have the same social desirability valence, that is, they should exclude activities that occur very frequently (>90%) or very infrequently (<10%) to ensure that respondents do not report to have engaged in all or none of the activities (Tsuchiya et al., 2007). Here, we selected inoffensive activities that integrate into the broader theme of participation. Our items are not directly related to turnout and radical right voting, but participation yet adverts to citizens' general civic duties. In addition, we chose inoffensive activities that are *fairly* common in Austria, but that citizens engage in more or less frequently: writing editorial letters, active participation in clubs or societies, changing the place of residence, and donating money. We summarize all other experimental conditions in Table 1. Detailed information about all surveys and the precise question wording is provided in the Supplementary Appendix.

Study A relies on a representative telephone survey conducted in 2011. Respondents were randomly assigned to a treatment ($n = 303$) and a control group ($n = 262$), where the control group was asked a short list of *four* items and the treatment group a long list of *five* items including whether they had voted in the 2008 election. As a modification to the traditional design, which was implemented in *Studies B* and *C*, we relied on Corstange's question format in *Study A* (Corstange, 2009). While the respondents in the treatment group are asked questions in the traditional question format—*how many* of the items on the list apply to them—Corstange suggests capturing the control group's

Table 1
The ICT's Properties (Studies A, B, and C)

	Study A		Study B		Study C	
Theme	Electoral turnout		Electoral turnout		Radical right-wing voting	
Mode	Telephone		Telephone		Online	
Response rate	50.3%[a]		37.4%[b]		—[c]	
	Control A	Treatment A	Control B	Treatment B	Control C	Treatment C
Question type	Which of . . .?	How many . . .?	How many . . .?	How many . . .?	How many . . .?	How many . . .?
List length	4	5	4	5	4	5
Response options	Wrote an editorial letter\ Was an active member in a club/society Moved to another city or community Donated money	Wrote an editorial letter Was an active member in a club/society Moved to another city or community Donated money Voted in the national parliamentary election in September 2008	Wrote an editorial letter Was an active member in a club/society Changed the main place of residence Donated money	Wrote an editorial letter Was an active member in a club/society Changed the main place of residence Donated money Voted in the national parliamentary election in September 2013	Wrote an editorial letter Was an active member in a club/society Changed the main place of residence Donated money	Wrote an editorial letter Was an active member in a club/society Changed the main place of residence Donated money Voted for the FPÖ in the national parliamentary election in September 2013
Sample size (*n*)	303	262	476	451	1,219	1,216

[a]Note that the field institute did not provide enough information to calculate AAPOR response rates as all "unknown others" are missing in the denominator.
[b]According to AAPOR Response Rate III.
[c]No response rate provided, as the online survey is based on a quota sample.

responses by asking them in *which* of the activities they have previously engaged. The benefit of this modification is that dichotomous variables are obtained for each response option in the control group, which allows estimating the probability of each item. On the basis of these probabilities, researchers are then enabled to also predict the probability of the sensitive behavior in the treatment group. While studies have shown that this technique permits further multivariate analysis based on the predicted probabilities, they also demonstrate that the question modification does not harm the core assumptions of the traditional ICT design (Corstange, 2009; Glynn, 2013).[6]

A postelection telephone survey was implemented in September 2013 (*Study B*) (Kritzinger, Thomas et al., 2016a; Kritzinger, Thomas, et al., 2016b). In comparison with 2011, we increased the sample sizes with 451 respondents participating in the treatment group and 476 in the control group. Both groups were asked the traditional ICT question *how many* of the list items they had engaged in. We used the same list length and items as in *Study A*.

Study C is based on a postelection panel survey that was conducted online on the basis of a quota sample (Kritzinger, Johann et al. 2016a; Kritzinger, Johann et al., 2016b). Instead of capturing turnout, we used the ICT to identify the proportion of respondents that voted for Austria's radical right party (*FPÖ*). The design allowed us to increase the sample size (treatment group: $n = 1{,}216$, control group: $n = 1{,}219$) again, but the list length and list items remained the same as in *Studies A* and *B*. Both groups were asked *how many* of the activities they had previously engaged in, the traditional ICT question format.

Direct self-reports of turnout[7] and vote choice were used to capture the proportion of voters and radical right supporters.[8] Following Holbrook and Krosnick (2010), we provide one-tailed z-tests to identify whether the ICT provides higher or lower estimates than the direct measures. We expect that the turnout rate as estimated by the ICT is significantly lower than self-reported turnout in *Studies A* and *B* (i.e., following the "less-is-better" assumption), and that the ICT estimate for radical right voting is significantly higher in *Study C* (i.e., following the "more-is-better" assumption).

Results

We begin by presenting our findings on socially desirable high-incidence behavior looking at voter turnout. While *Study A* yields an ICT estimate of 85.4%, self-report

[6]Consequently, we obtain dichotomous variables for each item on the control group's list. On the basis of these dummy variables, we then generated a simple count measure, and calculated the mean for the control group.

[7]We used the Belli, Moore, and VanHoewyk (2006) turnout question with additional response options.

[8]By contrast to the conditions implemented in the 2011 experiment, *all* respondents were also directly asked to provide direct self-reports in the surveys conducted in 2013. The literature has pointed out that question order or contamination effects, in which one question affects the response to a later question, may occur (Gaines, Kuklinski, & Quirk, 2007; Transue, Lee, & Aldrich, 2009). We reduced the possibility of potential spillover effects in 2013 by directly asking respondents to provide self-reports of turnout and vote choice at the beginning of the survey and by implementing the experiments by the end of the questionnaire. Respondents were asked to answer a large number of political and nonpolitical items in between. The fact that the ICT appears to perform poorly in 2011—where a separate experimental condition was implemented to capture the direct self-reports—and in 2013—where all respondents were directly asked to provide a direct self-report—suggests that spillover effects are not a problem in this context.

displays a turnout of 80.8%. That is a difference of 4.6% (z-statistic = 0.51, $p = 0.31$), which fails to reach conventional levels of statistical significance, however. Although the results indicate that the ICT does not perform differently in comparison with the direct measure, it is worth noting that its estimate is further away from the official turnout results than the traditional question format. In the 2008 general election, 78.8% of the eligible voters in Austria cast a ballot.[9] In sum, the ICT produces a decent estimate that is similar to the direct self-reports, but which is yet further away from the value provided by the official statistics than the traditional question format.

In *Study B*, the ICT yields a turnout estimate of 97.7% compared with the 92.7% of respondents who directly reported that they cast a ballot in 2013. Once again the difference between these two estimates is not statistically significant at conventional levels (z-statistic = 0.81, $p = 0.21$). However, the result also confirms our previous conclusions: In statistical terms, the ICT provides a similar result as the direct measure, but it is again further away from the official turnout result, which was declared at 74.9% in the 2013 elections.[10] Once again, we are unable to obtain a lower estimate in the logic of the "less-is-better" assumption[11]: We must conclude that the ICT failed to better predict turnout than the direct reports.

Next, we turn to underreporting of socially undesirable high-incidence behavior by looking at radical right voting (*Study C*). While 23.3% of the respondents in our panel study reported that they voted for the FPÖ looking at the traditional vote choice question, only 7.2% indicated that they supported the FPÖ according to the ICT's estimate. The z-statistics reveal a major statistically significant disparity between the ICT's estimate and the self-reported estimate (z-statistic = 3.4, $p < 0.01$). Once again, the ICT was unable to provide more accurate estimates in comparison with the direct self-reports. Considering that the share of FPÖ voters was declared by official government statistics with 20.5%,[12] the results based on the direct self-reports are closer to the official values than the ICT estimates. The findings suggest that the ICT also fails to limit underreporting of socially undesirable high-incidence behavior—the "more-is-better" assumption has to be rejected (Table 2).[13]

[9]See Bundesministerium für Inneres (2008). Österreichische Nationalratswahlen 2008 [Election Results]. Retrieved February 9, 2015 from http://wahl08.bmi.gv.at/.

[10]See Bundesministerium für Inneres (2013). Österreich–Nationalratswahl 2013 [Election Results]. Retrieved February 9, 2015 from http://wahl13.bmi.gv.at/.

[11]Although we use different research designs with regard to our list items (Corstange, 2009; Holbrook & Krosnick, 2010), we do not obtain better or worse ICT results regardless of the different formats. We think that this is a strength of our study and a contribution to the literature focusing on the various techniques to ask about items.

[12]See Bundesministerium für Inneres (2013). Österreich–Nationalratswahl 2013 [Election Results]. Retrieved February 9, 2015 from http://wahl13.bmi.gv.at/.

[13]Our approach to evaluating the ICT's performance implicitly relies on the assumption of *no* design effect, that is, the response behaviour is *not* dependent on the list length itself—independent from the item content. We have evaluated our three studies on the presence of such an effect (Blair & Imai, 2012, 2013; Comşa & Postelnicu, 2013; Wolter & Laier, 2014). While we do not find statistically significant support for potential design effects in Study B and Study C, we uncover a statistically significant result for Study A. The p-value in Study A is below the crucial threshold of $\alpha/2 = .025$ (Blair & Imai, 2012, p. 69). Consequently, we cannot completely rule out that the relatively poor performance of the ICT in Study A may at least in part reflect a design effect (Rosenfeld, Imai, & Shapiro, 2014). However, as the list items in Study A focus on different areas, we expect the design effect to be small (Comşa & Postelnicu, 2013).

Table 2
Estimates of the ICT Estimates and Self-Reports (Studies A, B, and C)[a]

	Study A: Turnout 2011	Study B: Turnout 2013	Study C: RRW 2013
Self-reports			
Sample size	260	935	2052
Self-report's estimate	80.8%	92.8%	23.3%
	(2.5%)	(0.9%)	(0.9%)
ICT			
Mean for treatment group	2.1	2.5	2.82
Mean for control group	1.3	1.6	2.75
Sample size	565	927	2435
ICT's estimate	85.4%	97.9%	7.2%
	(8.9%)	(6.3%)	(4.6%)
Official statistics	78.8%	74.9%	20.5%
z-test ICT and self-reports[b]	0.50	0.80	3.42***

*$p<.05$, **$p<.01$, ***$p<.001$. Standard errors in parentheses.
[a]Note that we also used the least squares estimator to calculate the ICT's results instead of the standard difference-in-means procedure (see Blair & Imai, 2012; Comşa & Postelnicu, 2013; Rosenfeld, Imai, & Shapiro, 2015; Çarkoğlu & Aytaç, 2015). The results do not substantially change our findings: Study A, 85.4% (8.9%); Study B, 98.0% (6.2%); Study C, 7.2% (4.6%). The maximum likelihood estimator did not converge (see Comşa & Postelnicu, 2013; Wolter & Laier, 2014 for a more detailed discussion).
[b]One-tailed z-tests. *$p < .05$, **$p < .01$, ***$p < .001$. Standard errors in parentheses. Note that standard errors of mean differences are reported for ICT estimates (see Holbrook and Krosnick, 2010). Data of *Study A* and *Study B* weighted.

Overall, the ICT has performed poorly looking at socially (un-)desirable *high*-incidence behavior.

Conclusion and Discussion

This article has investigated whether the ICT reduces SDB in high-incidence behavior. Previous evidence has indicated that the ICT is a useful method to reduce SDB when it is applied to socially undesirable low-incidence behavior. It is easily administered in questionnaires and allows respondents to confidentially reveal sensitive behavior. However, it is unclear whether the ICT also performs well in tackling SDB in socially desirable and undesirable high-incidence behavior. Our article contributes systematic empirical analysis of the latter in the context of electoral behavior. We tested whether the ICT is able to accurately predict socially desirable high-incidence behavior (*here*: electoral turnout) as well as socially undesirable high-incidence behavior (*here*: radical right voting). All results point into the same direction: When predicting high-incidence behavior, the ICT performs worse in comparison with the direct self-reports on all occasions. We further addressed some of the issues that were identified by previous studies, such as sample sizes, question format, and modes. However, changing these conditions did not improve the ICT's performance in the context of high-incidence behavior.

Our findings confirm doubts about the functioning of the ICT on socially (un-) desirable high-incidence behavior raised by previous research. This suggest that the ICT might be an appropriate method to capture socially undesirable, infrequently practiced behavior, such as drug abuse or shoplifting, but that it is inappropriate for predicting frequently practiced behavior with socially desirable and undesirable connotations, such as turnout and radical right voting.

We can only speculate about the reasons for this. As Holbrook and Krosnick (2010) have pointed out, other underlying mechanisms may be at work. The ICT's logic may simply not apply to high-incidence behavior. Social desirability pressures may also be much lower than previously expected when electoral behavior is concerned, indicating that overreporting occurs for other reasons than SDB. Although investigations using other variants of the randomized response technique suggest that such methods are valuable to capture electoral (non-)participation, they also indicate that voting is less sensitive in comparison with other behaviors, such as domestic violence (Moshagen, Musch, & Erdfelder, 2012). Some of the empirical properties of the ICT might render the design particularly vulnerable to error. One example is the selection of the unobtrusive items: The items in the Austrian context were carefully collected, but it is impossible to fully suspend that they had too much or too little prevalence. Finally, mode effects may play a role especially in Study C. Previous research has suggested that SDB is lower in online studies because respondents anonymously fill in the questionnaire without the presence of an interviewer (Holbrook & Krosnick, 2010). Moreover, online surveys attract young and politically interested respondents. As radical right voting is particularly prominent among younger citizens, the online mode may capture such effects.

Overall, we conclude that studies, which aim at exploring the psychological mechanisms of the ICT in high-incidence behavior, are necessary before the technique should be used in surveys again. This includes empirical research on the certainty with which we can identify that the high-incidence behavior in question is indeed prone to SDB, that is, electoral behavior in the respective context. It may well be that sensitive behavior, such as radical right support, is more accepted in other contexts, and other bias distorts the quality of the survey responses. We would advise researchers to be cautious when using the ICT in the context of socially (un-)desirable *high-incidence* behavior until these issues are solved. It may even be worth reconsidering more traditional question formats. At least our results lend some confidence to the assertion that staying away from ICT and exploring other methods to limit SDB when dealing with high-incidence behavior may be more valuable.

Supplementary Data

Supplementary Data are available at IJPOR online.

Funding

This work is supported by the Austrian National Election Study, a National Research Network sponsored by the Austrian Science Fund (S10902-G11).

Acknowledgments

The authors thank the anonymous reviewers for their constructive feedback and Graeme Blair and Kosuke Imai for their helpful support.

References

Ahart, A. M., & Sackett, P. R. (2004). A new method of examining relationships between individual difference measures and sensitive behavior criteria: Evaluating the unmatched count technique. *Organizational Research Methods*, *7*, 101–114. doi: 10.1177/1094428103259557

Aichholzer, J., Kritzinger, S., Wagner, M., & Zeglovits, E. (2014). How has radical right support transformed established political conflicts? The case of Austria. *West European Politics*, *37*, 113–137. doi: 10.1080/01402382.2013.814956

Belli, R. F., Moore, S. E., & VanHoewyk, J. (2006). An experimental comparison of question forms used to reduce vote overreporting. *Electoral Studies*, *25*, 751–759. doi: 10.1016/j.electstud.2006.01.001

Bernstein, R., Chadha, A., & Montjoy, R. (2001). Overreporting voting: Why it happens and why it matters. *Public Opinion Quarterly*, *65*, 22–44. doi: 10.2307/3078784

Biemer, P., & Brown, G. (2005). Model-based estimation of drug use prevalence using item count data. *Journal of Official Statistics*, *21*, 287–308. doi:10.1016/j.electstud.2006.01.001

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

Blais, A. (2010). Political participation. In L. LeDuc, R. G. Niemi, & P. Norris (Eds.), *Comparing democracies 3. Elections and voting in the 21st century* (pp. 165–183). Thousand Oaks, CA: Sage.

Çarkoğlu, A., & Aytaç, S. E. (2015) Who gets targeted for vote-buying? Evidence from an augmented list experiment in Turkey. *European Political Science Review*, *7*, 547–566. doi:10.1017/S1755773914000320

Coms¸a, M., & Postelnicu, C. (2013). Measuring social desirability effects on self-reported turnout using the Item Count Technique. *International Journal of Public Opinion Research*, *25*, 153–172. doi: 10.1093/ijpor/eds019

Corstange, D. (2009). Sensitive questions, truthful answers? Modeling the list experiment with LISTIT. *Political Analysis*, *17*, 45–63. doi: 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 and Research*, *40*, 169–193. doi: 10.1177/0049124110390768

Dalton, D. R., Wimbush, J. C., & Daily, C. M. (1994). Using the Unmatched Count Technique (UCT) to estimate base rates for sensitive behavior. *Personnel Psychology*, *47*, 817–829. doi: 10.1111/j.1744-6570.1994.tb01578.x

de Jonge, C. P. K., & Nickerson, D.W. (2014). Awrtificial inflation or deflation? Assessing the Item Count Technique in comparative surveys. *Political Behavior*, *36*, 659–682. doi: 10.1007/s11109-013-9249-x

Diaz-Cayeros, A., Magaloni, B., Matanock, A., & Romero, V. (2011). Living in fear: Social penetration of criminal organizations in Mexico. Unpublished paper. Retrieved from http://irps.ucsd.edu/assets/001/502967.pdf. Retrieved April 14, 2014

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. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 185–210). New York, NY: Wiley. doi: 10.1002/9781118150382.ch11

Flavin, P., & Michael Keane, M. (2009). How angry am I? Let me count the ways: Question format bias in list experiments. Unpublished paper. Retrieved from http://files.campus.edublogs.org/blogs.baylor.edu/dist/2/1297/files/2010/09/List_Experiment-1zftaac.pdf. Retrieved May 7, 2014.

Gaines, B. J., Kuklinski, J. H., & Quirk, P. J. (2007). The logic of the survey experiment reexamined. *Political Analysis*, *15*, 1–20. doi: 10.1093/pan/mpl008

Gilens, M., Sniderman, P. M., & Kuklinski, J. H. (1998). Affirmative action and the politics of realignment. *British Journal of Political Science*, *28*, 159–183.

Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly*, *77*(Suppl 1), 159–172. doi: 10.1093/poq/nfs070

Gonzalez-Ocantos, E., de Jonge, C. D., Meléndez, C., Osorio, J., & Nickerson, D. W. (2012). Vote buying and social desirability bias: Experimental evidence from Nicaragua. *American Journal of Political Science*, *56*, 202–217. doi: 10.1111/j.1540-5907.2011.00540.x

Heerwig, J. A., & McCabe, B. J. (2009). Education and social desirability bias: The case of a black presidential candidate. *Social Science Quarterly*, *90*, 674–686. doi: 10.1111/j.1540-6237.2009.00637.x

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

Imai, K. (2011). Multivariate regression analysis for the Item Count Technique. *Journal of the American Statistical Association*, *106*, 407–416. doi: 10.1198/jasa.2011.ap10415

Janus, A. L. (2010). The influence of social desirability pressures on expressed immigration attitudes. *Social Science Quarterly*, *91*, 928–946. doi: 10.1111/j.1540-6237.2010.00742.x

Johann, D., Thomas, K., Faas, T., & Fietkau S. (2015). Alternative Messverfahren rechtpopulistischen Wählens im Vergleich: Empirische Erkenntnisse aus Deutschland und Österreich. In H. Schoen & B. Weßels (Eds.), *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2013*, pp. 447–470, Wiesbaden: VS Verlag.

Kane, J. G., Craig, S. C., & Wald, K. D. (2004). Religion and presidential politics in Florida: A list experiment. *Social Science Quarterly*, *85*, 281–293. doi: 10.1111/j.0038-4941.2004.08502004.x

Karp, J. A., & Brockington, D. (2005). Social desirability and response validity: A comparative analysis of overreporting voter turnout in five countries. *Journal of Politics*, *67*, 825–840. doi: 10.1111/j.1468-2508.2005.00341.x

Kritzinger, S., Johann, D., Glantschnigg, C., Aichholzer, J., Glinitzer, K., Thomas, K., … & Zeglovits, E. (2016a). *AUTNES TV debate panel study 2013*. GESIS Data Archive, Cologne. ZA5858 Data File Version 2.0.1. doi:10.4232/1.12450.

Kritzinger, S., Johann, D., Glantschnigg, C., Aichholzer, J., Glinitzer, K., Thomas, K., … & Zeglovits, E. (2016b). *AUTNES TV debates panel study 2013 – documentation*. Vienna: University of Vienna.

Kritzinger, S., Thomas, K., Glantschnigg, C., Aichholzer, J., Glinitzer, K., Johann, D., … & Zeglovits, E. (2016a). *AUTNES Comparative Study of Electoral Systems post-election survey 2013*. (www.autnes.at). GESIS Data Archive, Cologne. ZA5856 Data File Version 2.0.0. doi:10.4232/1.12423.

Kritzinger, S., Thomas, K., Glantschnigg, C., Aichholzer, J., Glinitzer, K., Johann, D., … & Zeglovits, E. (2016b). *AUTNES Comparative Study of Electoral Systems post-election survey 2013 – documentation*. Vienna: University of Vienna.

Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial attitudes and the "new south". *The Journal of Politics*, *59*, 323–349. doi: 10.2307/2998167.

LaBrie, J. W., & Earleywine, M. (2000). Sexual risk behaviors and alcohol: Higher base rates revealed using the Unmatched-Count Technique. *The Journal of Sex Research*, *37*, 321–326. doi: 10.1080/00224490009552054

Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M., & Maas, C. J. M. (2005) Meta-analysis of randomized response research. Thirty-five years of validation. *Sociological Methods & Research*, *33*, 319–348. doi: 10.1177/0049124104268664

Martinez, M. D., & Craig, S. C. (2010). Race and 2008 presidential politics in Florida: A list experiment. *The Forum*, *8*, 1–14. doi: 10.2202/1540-8884.1316

Moshagen, M., Hilbig, B. E., Erdfelder, E., & Moritz, A. (2013). An experimental validation method for questioning techniques that assess sensitive issues. *Experimental Psychology 61*, 48–54. doi: 10.1027/1618-3169/a000226

Moshagen, M., Musch, J., & Erdfelder, E. (2012). A stochastic lie detector. *Behavior Research Methods*, *44*, 222–231. doi: 10.3758/s13428-011-0144-2

Nepusz, T., Petróczi, A., Naughton, D. P., Epton, T., & Norman, P. (2014). Estimating the prevalence of socially sensitive behaviors: Attributing guilty and innocent noncompliance with the Single Sample Count method. *Psychological Methods*, *19*, 334–355. doi: 10.1037/a0034961

Petróczi, A., Nepusz, T., Cross, P., Taft, H., Shah, S., Deshmukh, N. … & Naughton, D. P. (2011). New non-randomised model to assess the prevalence of discriminating behaviour: A pilot study on mephedrone. Substance Abuse Treatment, Prevention, and Policy, *6*, 1–20. doi:10.1186/1747-597X-6-20

Phillips, D. L., & Clancy, K. J. (1972). Some effects of "social desirability" in survey studies. *American Journal of Sociology*, *77*, 921–940. doi: 10.2307/2776929

Rayburn, N. R., Earleywine, M., & Davison, G. C. (2003). An investigation of base rates of anti-gay hate crimes using the unmatched-count technique. *Journal of Aggression, Maltreatment & Trauma*, *6*, 137–152. doi: 10.1300/J146v06n02_07

Redlawsk, D. P., Tolbert, C. J., & Franko, W. (2010). Voters, emotions, and race in 2008: Obama as the first black president. *Political Research Quarterly*, *63*, 875–889. doi 10.1177/1065912910373554

Rosenfeld, B., Imai, K., & Shapiro, J. (2015). An empirical validation study of popular survey methodologies for sensitive questions. *American Journal of Political Science*. doi: 10.1111/ajps.12205

Silver, B. D., Anderson, B. A., & Abramson, P. R. (1986). Who overreports voting? *American Political Science Review*, *80*, 613–624 doi: 10.1086/269082

Streb, M. J., Burrell, B., Frederick, B., & Genovese, M. A. (2008). Social desirability effects and support for a female American president. *Public Opinion Quarterly*, *72*, 76–89. doi: 10.1093/poq/nfm035

Transue, J. E., Lee, D. J., & Aldrich, J. H. (2009). Treatment spillover effects across survey experiments. *Political Analysis*, *17*, 143–161. doi: 10.1093/pan/mpn012

Traugott, M. W., & Katosh, J. P. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly*, *43*, 359–377. doi: 10.1086/268527

Tsuchiya, T. (2005). Domain estimators for the item count technique. *Survey Methodology*, *31*, 41–51.

Tsuchiya, T., Hirai, Y., & Ono, S. (2007). A study of the properties of the item count technique. *Public Opinion Quarterly*, *71*, 253–272. doi: 10.1093/poq/nfm012

Umesh, U. N., & Peterson, R. A. (1991). A critical evaluation of the Randomized Response Method – applications, validation, and research agenda. *Sociological Methods and Research*, *20*, 104–138. doi: 10.1177/0049124191020001004

Werts, H., Scheepers, P., & Lubbers, M. (2013). Euro-scepticism and radical right-wing voting in Europe, 2002-2008: Social cleavages, socio-political attitudes and contextual characteristics determining voting for the radical right. *European Union Politics*, *14*,183–205. doi: 10.1177/1465116512469287

Wolter, F., & Laier, B. (2014). The effectiveness of the Item Count Technique in eliciting valid answers to sensitive questions: An evaluation in the context of self-reported delinquency. *Survey Research Methods*, *8*, 153–168.

Zeglovits, E., & Kritzinger, S. (2014). New attempts to reduce overreporting of voter turnout and their effects. *International Journal of Public Opinion Research*, *26*, 224–234. doi: 10.1093/ijpor/edt010

Biographical Notes

Kathrin Thomas is Research Associate in the Centre for Comparative Social Surveys at the City University London and Research Fellow at the University of Vienna. She holds a PhD in Politics from the University of Exeter.

David Johann is Researcher at the German Centre for Research on Higher Education and Science Studies (DZHW) and Senior Research Fellow at the University of Vienna. David Johann holds a PhD in Political Science from the University of Vienna.

Sylvia Kritzinger is Professor and Head of the Department of Methods in the Social Sciences at the University of Vienna. She is one of the principle investigators of the Austrian National Election Study. Sylvia Kritzinger holds a PhD in Political Science from the University of Vienna.

Carolina Plescia is Assistant Professor in the Department of Methods in the Social Sciences at the University of Vienna. She holds a PhD in Political Science from Trinity College Dublin, Ireland.

Eva Zeglovits is Managing Director of the Institut für Empirische Sozialforschung (IFES), a private research company in polling and applied research. She is also Research Fellow at the University of Vienna. Eva Zeglovits holds a PhD in Political Science from the University of Vienna.