Erschienen in: Devianz und Subkulturen : Theorien, Methoden und empirische Befunde / Krumpal, Ivar; Berger, Roger (Hrsg.). - Wiesbaden : Springer VS, 2020. - S. 123-149. - ISBN 978-3-658-27227-2
https://dx.doi.org/10.1007/978-3-658-27228-9_5

Let's Ask About Sex: Methodological Merits of the Sealed Envelope Technique in Face-to-Face Interviews

Felix Wolter and Peter Preisendörfer

When it comes to deviant behavior and other sensitive topics, respondents in surveys tend either to refuse answering such sensitive questions or to tailor their answers in a socially desirable direction. This occurs by underreporting on negatively connoted behaviors or attitudes and by overreporting on positively connoted ones (be it deliberately or undeliberate). Thus, prevalence estimates of deviant traits are biased. Moreover, if the tendency to misreport is related to influencing factors of the deviant behavior or attitude under investigation, the correlations are biased as well.

In order to tackle the problems of nonresponse and misreporting to sensitive questions, survey methodology has come up with several special questioning

This research was conducted within the projects "Asking Sensitive Questions: Possibilities and Limits of Randomized Response and Other Techniques in Different Survey Modes", and "Sensitive Questions and Social Desirability: Theory and Methods", both supported by the German Research Foundation DFG (Priority Program SPP 1292, Grant PR 237/6-1 and Grant WO 2242/1-1).

F. Wolter (✉) · P. Preisendörfer
Johannes Gutenberg-Universität Mainz, Mainz, Germany
E-Mail: felix.wolter@uni-mainz.de

P. Preisendörfer
E-Mail: preisendoerfer@uni-mainz.de

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-2-wd8li1hm0lzy2

techniques which mostly aim at anonymizing the interview situation and in doing so, to impose less burden on respondents when answering sensitive questions. Among the oldest proposed techniques is the sealed envelope technique (SET) for face-to-face interviews: Respondents are asked to fill out the "sensitive part" of the interview on a self-administered ballot and to seal it in a secret envelope that the interviewer does not open personally.

The aim of our paper is to investigate whether SET has methodological advantages as compared to classic direct questioning (DQ). The data stem from a "sensitive topic survey" among University students in the City of Mainz (Germany, n = 578). The sensitive questions selected for this evaluation pertain to sexual experiences and behavior.

Results show that—compared to direct questions—the sealed envelope technique has some advantages: It reduces misreporting of sensitive behavior; in particular, it helps certain subgroups (e.g. religious people) to overcome subjective barriers in answering sensitive questions accurately; it diminishes item nonresponse; and it lowers subjective feelings of uneasiness of interviewers and respondents when it comes to sensitive issues. The general conclusion is that the sealed envelope technique seems to be a helpful tool in gathering less biased data about sensitive behavior.

Survey methodology · Sensitive questions · Social desirability · Response bias · Sealed envelope technique · Sexual behavior

Umfragemethoden · Heikle Fragen · Soziale Erwünschtheit · Antwortverzerrung · Sealed envelope technique · Technik des verschlossenen Kuverts · Sexualverhalten

1 Introduction

Research on deviant behaviors or attitudes with survey data often raises problems of measurement validity. Since deviant traits, to varying degrees, conflict with existing social norms and/or pertain to private issues, many respondents have reservations about disclosing personal information about them in survey interviews. Consequently, they tend either to refuse answering such sensitive questions at all or to tailor their answers in a socially desirable direction. This occurs by underreporting

on negatively connoted behaviors or attitudes and by overreporting on positively connoted ones (be it deliberately or undeliberate). Thus, prevalence estimates of deviant traits are biased by the degree of response bias. Moreover, if the tendency to misreport is related to factors influencing the deviant behavior or attitude under investigation, the correlations between these influence factors and the sensitive items are biased as well.

Speaking more generally, sensitive questions in surveys pertain to personal, secret, socially frowned upon or illegal behavior, attitudes or personal characteristics. Obtaining valid answers to such questions is still a difficult endeavor, although numerous studies have explored possibilities of "Asking the Embarrassing Question" (Barton 1958) since the very beginning of survey research. Lee (1993, p. 3) notes that the difficulties generated by sensitive topics "tax the methodological ingenuity of social scientists" and "have produced a range of methodological innovations." Studies on how best to ask sensitive questions are therefore valuable for survey methodology in general in that they can stimulate creative approaches to other survey problems and provide evidence on the effectiveness of new survey techniques.

Looking at the literature about sensitive questions in surveys (Lee 1993; Barnett 1998; Tourangeau and Yan 2007; Krumpal 2013; Wolter 2012; Jann et al. 2019), we can see that the research agenda is characterized by two areas of discussion: (1) The first area deals with conceptual and theoretical issues of research about sensitive topics. Typical questions include: What are sensitive questions, how can we define them? Which attributes of questions and answers are responsible for the perception of sensitivity? Under what conditions and to what extent can we expect response biases? Are there changes over time in the degree of sensitivity of certain topics, and differences between (sub–) cultures? (2) A second line of research focuses on specific methods and techniques of how to obtain more accurate, less distorted answers to sensitive questions—compared to direct questioning. These techniques range from wording and framing techniques (e.g., Cantril 1940; Tourangeau and Smith 1996); sealed envelopes (Perry 1979) and vignettes (Auspurg and Hinz 2014) to randomized response (Fox and Tracy 1986) and item count techniques (Droitcour et al. 1991; Wolter and Laier 2014). Of course, these techniques can be used in different survey modes (face-to-face, telephone, mail, internet), resulting in mode effects.

Embedded in this broader field of research, the present article has a restricted scope. It is confined to one technique in one mode, namely the sealed envelope technique (SET) in face-to-face interviews. Far from being new, SET is a fairly traditional technique of collecting sensitive data (Barton 1958; Sudman and Bradburn 1974, 1982; Perry 1979; De Leeuw 2001; Krosnick et al. 2002). Nevertheless (and in contrast to other procedures such as wording/framing or

randomized response), detailed methodological reports about experiences with this technique are surprisingly rare (see, however, Makkai and McAllister 1992; Becker and Günther 2004; Becker 2006; Benson 1941; Aquilino 1998; Rasinski et al. 1999). Based on this deficit, it seems worthwhile to enlarge our knowledge about the possible merits of SET.

The usual procedure in applying SET is as follows: When it comes to one or more sensitive topics in the course of a face-to-face interview, the interviewer stops and begins to explain that based on the content of the questionnaire a particular problem of confidentiality may arise at this point. To guarantee anonymity, the questions will not be asked and answered orally but in writing. The respondents are given a separate sheet of paper or a separate short questionnaire, asked to fill it out by themselves, and then seal it in an envelope which only the final researcher is allowed to open, not the interviewer.

The general idea behind this procedure is to avoid invalid answers by the interviewee caused by the direct interaction with the interviewer. Based on a rational choice perspective of survey response behavior (see, e.g., Esser 1991; Tourangeau et al. 2000, pp. 281–284; Stocké 2007b, 2004; Wolter 2012), we can expect respondents to decide answering questions in a way that reduces potential costs, threats and risks to themselves. In the case of sensitive questions, such costs and threats encompass invasion of privacy, disclosure to third parties, and—probably most importantly—disapproval from the interviewer (Lee 1993, pp. 3–11; Tourangeau et al. 2000, pp. 257–259). Respondents are interested in making a good impression to the interviewer and in avoiding answers possibly inducing the image of being a "bad guy". Consequently, they do not lie "just for fun," but in a way and moving in a direction that can be predicted by subjective expected utility theory.[1] Assuming that answering survey questions is a goal-directed, instrumentally rational selection between response options, the overall prediction is that interviewees tend to underreport socially undesirable behaviors and to overreport socially expected ones. Not only the respondent, but also the interviewer is interested in a smooth interview relationship. Sensitive questions therefore present a similar threat to the interviewer. Empirical studies have shown that response distortions for sensitive questions often have more to do with the interviewer's discomfort or worries that asking the question will be problematic than with discomfort of the respondent (Bradburn and Sudman 1979, chap. 4; Schnell

[1]With respect to social desirability bias, rational choice theory focuses on three determinants of this type of misreporting (Stocké 2007b): 1) respondents' social desirability (SD) beliefs, i.e. their perceptions of which answer is socially expected, 2) respondents' need for social approval, and 3) respondents' feeling of privacy in the interview situation.

1997, pp. 274–275; Hox and De Leeuw 2002; Biemer and Lyberg 2003, p. 180; Schnell and Kreuter 2005). The baseline assumption behind SET is that its application makes it easier both for the interviewer and the interviewee, because both parties do not have to speak about and openly interact about socially undesirable behavior. This results in the working hypothesis: The more threatening—for the total population or for certain subgroups of respondents—a question is, the more useful in comparison to direct questioning SET will be.

To examine this hypothesis, the present article refers to sexual behaviors such as coital frequency or masturbation, a classic field of sensitive topic research (Tourangeau et al. 1997; Tourangeau and Smith 1996). At least some of these behaviors can be judged as potentially deviant acts, e.g., a very high or low number of lifetime sexual partners, or sexual infidelity. Based on a survey with 578 students of the University of Mainz (Germany), SET answers will be compared to direct questioning (DQ) answers, i.e. answers given in response to conventional direct questions. The main criterion for the comparison and for the evaluation of SET is the presumed under- or overreporting bias. To what extent is SET successful in alleviating misreporting associated with direct questions? The response bias problem will be addressed using simple descriptive as well as multivariate analyses. In doing so, we rely on the "more (less) is better" assumption, meaning that for socially undesirable (desirable) behaviors, higher (lower) prevalence estimates are considered being more valid. Besides the under- or overreporting criterion, some additional and supplementary criteria will be used to evaluate the merits of SET, namely item-nonresponse rates, respondents' subjectively perceived uneasiness in answering the sex items, and—on a qualitative basis—the interviewers' discomfort when posing these questions.

Section 2 presents the study design and variables of the empirical analyses. Sect. 3 first presents the descriptive findings with respect to the prevalence of different sexual behaviors in a mode-comparing perspective (DQ vs. SET). Second and using multivariate regression analyses, we investigate whether factors influencing the sexual behaviors are different for the two question modes. Sect. 4 will address results regarding nonresponse and respondents' and interviewers' uneasiness. The article ends with a short conclusion in Sect. 5.

2 Study Design

The data used in this article stem from a teaching research project in a one-year seminar on social research methods. In October and November 2007, a group of 578 students of the University of Mainz were interviewed by 47 seminar participants. The interviews were face-to-face, and the average interview duration was

about 30 minutes. The 578 interviews resulted from a random sample of 1,508 addresses drawn from the student registration office of the University of Mainz. About 34,000 students were registered at the University of Mainz in 2007. Not all students, however, had a chance of being included in the sample because we imposed two restrictions: The students selected by the registration office had to be "regular students" (not short-term guest students, doctoral students, etc.) and students living in the City of Mainz (not students from outside, i.e. from other towns and villages in the Rhine/Main metropolitan area).[2] Each of the 47 seminar participants received 30 addresses, and the instruction was that s/he should complete 15 interviews. The amount of work put in and the success of the student interviewers varied: some did not use all their addresses, others required additional addresses, and the majority completed less than the target number of 15 interviews. When we ignore unused addresses and other addresses not belonging to the target population (e.g. university dropouts), the overall response rate is 43%.

In the initial step, the students selected received a personal letter informing them about the study, asking them to participate, and announcing the visit of an interviewer (whose name, phone number and e-mail address were provided). This invitation letter introduced the survey under the heading "Campus and Beyond" and explained that student life on campus and in the City of Mainz will be the content of the interview. It also cautiously mentioned that the questionnaire will touch sensitive topics "which people often don't like to talk about." In fact, the interview could be categorized as a "sensitive topic survey", i.e. most question modules focused on sensitive topics. It is possible that this reference to sensitive questions induced some students to refuse to participate, and hence caused a self-selection bias. It can be assumed, however, that this bias is of restricted relevance, because students are a population group which is relatively open and tolerant with respect to discussing controversial topics, including sensitive ones.

There were two versions of the questionnaire: In the DQ version, all sensitive questions were asked directly. In the "indirect version", different techniques to ask several sensitive topics were applied. Besides wording/framing techniques

[2]A third restriction imposed by the registration office was that the students selected were born between May and October. The interviewers did not know about this restriction and asked for the month (and year) of birth in their interviews. This enabled a check of interviewer reliability. Only three interviewers reported birth months outside the May-October range. The field work of these three interviewers was checked thoroughly, but without confirmation of serious cheating.

(used for questions about drug abuse and petty criminal behavior) and the randomized response technique (used to gather information about cheating in the university system), the indirect version used SET to investigate sexual experiences and behaviors of the respondents. As mentioned above, the SET results will be the subject of this paper.[3] The overall expectation was that for socially undesirable (desirable) traits, the indirect SET version would result in higher (lower) prevalence estimates than the DQ version. With the initial objective of having one third direct and two thirds indirect interviews, the interviewers were randomly assigned to conduct direct or indirect interviews. The final outcome was that 211 (37%) of all 578 interviews were direct interviews and 367 (63%) indirect ones.[4] Concerning response rates (unit nonresponse), no significant differences could be observed between direct and indirect interviews.

Even though most questions were sensitive, almost all respondents successfully finished the interview (only 4 break-offs). As in most surveys, the crucial barriers of participation were first that the addressees were not at home and could never be reached, and secondly that they refused to participate at all, i.e. did not even start the interview. With respect to the 578 final respondents, the interviewers had the impression that their willingness to cooperate was "high" in 90% of cases, "moderate" in 9% and "low" in 1% of cases. Furthermore, on a 5-digit scale (ranging from 1 = completely unreliable to 5 = completely reliable) the interviewers rated the answers of 89% of the respondents as "reliable" or "completely reliable."

The "sex module" of the questionnaire was located near the end of the interview (prior to the closing socio-demographic questions). It covered 20 questions, but only seven of them can be qualified as core sensitive questions. The other questions served as an introduction, were tailored to restricted subgroups, or had other special purposes in the interview process.[5] The interviewers of the SET

[3]Wording/framing techniques did not make much difference to direct questioning in general; the randomized response procedure used in our survey proved to deliver inconsistent and implausible results (for descriptive findings, see Preisendörfer 2008).

[4]Each of the 47 interviewers had an interviewer number. The randomization procedure used this number to determine which type of interview (direct/indirect) an interviewer had to conduct. This assignment of interviewers to questionnaire versions includes the risk to confound interviewer attributes and treatment (direct/indirect), but the randomization should control for this.

[5]For example, questions on whether respondents live in a regular partnership or what they consider important for living in a partnership ("love", "fun", "freedom", etc.).

application had an envelope marked "Confidential." Inside this envelope was a short questionnaire containing the questions on sexuality issues. The interviewers handed this envelope to the respondent and gave the following instructions: "May I ask you to take the separate short questionnaire out of this envelope, to fill out this questionnaire by yourself, and to put it back into the envelope. Then, you can close and seal the envelope. I am not allowed to open it; this can only be done by a research fellow at the sociology institute who doesn't know you. This arrangement guarantees your anonymity, i.e. all your answers will remain strictly confidential." Table 1 depicts the question wording of the seven sex questions and a question on respondents' uneasiness which was posed at the end of the interview. In the DQ and SET version the questions were identical in layout and wording.

Table 1 Question Wordings: Dependent Variables

Item	**Wording**
Sexual partners (lifetime)	"How many different sexual partners have you had in your lifetime so far? If you do not remember this exactly, please try to give an estimation."
Age of first sex	"At which age did you have sex for the first time?"
Coital frequency (last 4 weeks)	"In the last four weeks, how many times have you had sexual intercourse? If you do not remember this exactly, please try to give an estimation."
One-night stand (ever)	"Concerning your sexual experiences, have you ever had a one-night stand?" (with answers yes/no)
Sexual infidelity in current/last relationship (ever)	"In the time of your current (last) partnership, have you ever had an outside sexual affair? An outside sexual affair means that you had sex with a person other than your current (last) life partner" (with answers yes/no)
Homosexual contact (ever)	"Have you ever had a homosexual sexual contact?" (with answers yes/no)
Masturbation (often, very often)	"Although scientific studies show that more than 80 percent of all men and more than 60 percent of all women masturbate, most people don't like to talk about this. Nevertheless, we would like to know: How often do you masturbate?" (with answers very often, often, sometimes, seldom, never)
Respondent uneasiness	"How uneasy did you personally feel in answering our questions about sexual experiences and behavior?" (with a 10-digit scale from 1 = not uneasy at all to 10 = very uneasy)

In our multivariate analyses, effects of some relevant covariates on the sexual behaviors will be investigated. These independent variables and their distributions are given in Table 2. Gender, age, and partnership (whether the respondent currently has a regular life partner) have been selected as independent variables because they constitute indispensable basics of all research about sexual behavior (e.g., Jasso 1985; Tourangeau et al. 1997). Religious affiliation is intended to capture a conservative worldview. Because many religions address sexual issues and impose—to varying degrees—explicit and often strict norms regarding sexual behavior, a strong religious affiliation might both affect the true sexual behavior of people and raise reservations of telling about sexuality and/or reporting negatively connoted sexual behavior in survey interviews. The last indicator, partying habits, are seen as a mixed indicator of personal needs as well as situational opportunities for sexual experiences. The distributions by question format in Table 2 show some (marginally) significant differences for age, partnership, and religiousness. Hence, the randomization did not work perfectly well; the amount of the differences, however, is not very strong.

Table 2 Independent Variables

	DQ		SET		Diff: χ^2 or t	n
	Mean	SD	Mean	SD		
Gender (1 = female)	57.8		58.9		0.06	578
Age (years)	24.6	3.6	25.4	3.9	2.23*	575
Partnership (1 = yes)	53.6		61.0		3.04+	575
Religiousness [1…5]	2.3	1.2	2.5	1.2	1.69+	575
Party affinity [1…5]	2.6	0.9	2.5	1.0	0.99	576

Depicted are mean values and standard deviations for metric items and percent values for binary ones. Differences by question mode were tested for by using t-tests for metric items and χ^2-tests for binary items. Question wording of (1) religiousness: "To which degree do you feel religious, that is, connected to religious ideas and religious bodies of thought?" with answers very strong, strong, partly, little, not at all (answers recoded into 1 = not at all to 5 = very strong). (2) Party affinity: "How often do you go to university parties?" with answers very often, often, sometimes, rarely, never (answers recoded into 1 = never to 5 = very often).
DQ = direct questioning, SET = sealed envelope technique.
*+: $p<0.1$; *: $p<0.05$*

3 Results 1: Prevalence Estimates and Determinants of Sexual Behaviors by Question Mode

Table 3 shows prevalence estimates and inference statistics for the seven sexual behavior items by question mode DQ versus SET. Concerning the number of sexual partners in their lifetime to date, no significant difference can be observed between DQ and SET with respect to the mean values. Nevertheless, an inspection of the frequency distributions (not documented) reveals two interesting details: (1) The slightly higher mean in the SET version is mainly caused by two outliers reporting 100 sexual partners. It may be that the anonymity of the sealed envelope prompted these two persons (one man and one woman) to have a joke, but it may also be that the SET made it easier to report such promiscuity.[6] (2) The above-mentioned finding stimulated the hypothesis that there is an inverted u-shaped relation between the number of sexual partners and social desirability (SD). Whereas a low and a high number of sexual partners are "sensitive answers," a moderate number fits social norms and expectations (for further empirical studies on such non-monotonic, inverted u-shaped patterns of social desirability beliefs, see Stocké and Hunkler 2007). This hypothesis is indirectly supported by comparing the DQ and SET distributions: Recoding the number of sexual partners pragmatically as low (0–2 partners), moderate (3–6 partners) and high (7 or more partners) yields a DQ distribution of 31%, 43%, and 26%, and a SET distribution of 35%, 34%, and 31% ($\chi^2 = 4.8$, $p = 0.091$). Hence, the moderate answer is clearly less pronounced in the SET version, and this allows the interpretation that SET seems to reduce the tendency to give answers in accordance with social desirability. This is furthermore confirmed if we generate a dummy variable indicating whether a very low (0–2) or high (>6) number of sexual partner was reported: This supposedly negatively connoted behavior is significantly more often reported in SET mode (see Table 3), with a difference of about nine percentage points.

The item "age of first sex" shows a significant difference between DQ and SET. In the latter group, respondents report having had their first sex at nearly one year later than in the DQ group. This finding is robust also if some outliers

[6]From further inspection of the two cases it is difficult to tell whether the reported 100 sex partners are plausible values, misreporting, or a coding error.

Table 3 Prevalence Estimates of Sexual Behaviors by Question Mode

		DQ (n = 211)	SET (n = 367)	Diff.: t or χ^2
Sexual partners (lifetime)	Median:	4.0	4.0	
	Mean:	5.9	6.6	0.94
	Low or high:	56.9	66.2	4.78*
Age of first sex	Median:	17	17	
	Mean:	17.0	17.9	4.06***
	>17 years:	31.2	48.8	16.26***
Coital frequency (last 4 weeks)	Median:	3.0	3.0	
	Mean:	4.7	5.0	0.51
	Low or high:	57.2	54.6	0.36
One-night stand (ever)		53.8	49.3	1.09
Sexual infidelity in current/last relationship (ever)		12.5	10.5	0.49
Homosexual contact (ever)		14.8	18.5	1.25
Masturbation	Median:	3.0	3.0	
	Mean:	2.6	3.0	3.52***
	Often, very often:	14.6	29.4	15.38***
Global test: all items		34.4	39.8	3.44***

Depicted are median and mean values for metric items and percent values for binary ones. Differences by question mode were tested for by using t-tests for metric items and χ^2-tests for binary items. See the main text body for an explanation of the global test.
DQ direct questioning, SET sealed envelope technique.
**: $p < 0.05$; ***: $p < 0.001$*

are deleted. If we compare the prevalence of students reporting having had no intercourse before the age of 18, the difference is even more striking: While in DQ mode, only 31% of the respondents say so, 49% report this in SET mode.

No significant differences between DQ and SET can be found in Table 3 for the four items "coital frequency in the last four weeks," "one-night stand," "sexual infidelity" in the current or, if there was no current partnership, in the last partnership and "homosexual contact." Although the two maxima of coital frequency (27 and 30) are reported in the SET version (one obviously wrong value of 300 reported intercourses within the last four weeks was deleted), the means are very similar. Furthermore, the additional finding (2) for the number of sexual

partners does not hold for coital frequency. About one third of the students interviewed did not have any intercourse in the last four weeks, mainly because they did not live in a partnership.[7] Half of the students report one-night stands, over 10% confess sexual infidelity in partnerships, and around 15% can look back at homosexual experiences.

Finally, and significantly, DQ and SET yield percentages of masturbation which clearly differ. In DQ mode, 14.6% admit to masturbating often or very often, as compared to 29.4% in SET mode, i.e. the percentage value doubles. This can be qualified as an important advantage of SET, because prior empirical research on the degree of sensitivity of different behaviors shows that masturbation is a highly sensitive topic. In an often-quoted sensitivity "hitlist" by Bradburn and Sudman (1979, p. 17), for example, masturbation ranks first. The Bradburn/Sudman list refers to the US population and is rather old, but we are convinced that the top rank of masturbation also applies to our sample of German students. Indeed, for the other items in Table 3 (from sexual partners to homosexual contacts) it could be argued that their sensitivity is lower, nonlinear (as for the number of sexual partners) or even that these are not at all sensitive. Masturbation, however, is undoubtedly a sensitive issue (see again Bradburn and Sudman 1979; Tourangeau and Yan 2007 for more data supporting our claim).

We furthermore conducted a global test by testing all seven items at once for mode differences. For the variables "lifetime sexual partners", "age of first sex", and "coital frequency", the dummy variables (see Table 3) were used for this purpose. Hence, all variables were coded in a way that indicates the assumed socially undesirable answer (e.g., having had a very low or high number of sexual partners and coital frequency etc.). The global test was carried out by stacking the data into long format and adjusting the standard errors for the clustering by respondent (using the Stata routine suest). The test yields a mean DQ-SET difference of 5.4% points, which is significant. Hence, we conclude that the overall result is that SET is helpful in alleviating underreporting to sensitive questions on sexual behavior.

In a second step, we conducted multivariate regression models with the seven sexual behaviors as dependent variables and the above introduced respondent characteristics as independent variables. First, we saw from Table 2 that the randomization between experimental groups did not work perfectly well; hence, we are going to look at the SET effect after controlling for other variables. Second,

[7]Confined to those living in a partnership, the mean of coital frequency in the last four weeks goes up to 8, and this corresponds with values obtained from other studies of coital frequency in partnership relations (e.g., Jasso 1985).

these other determinants of the respective sexual behaviors may be of different strength in the SET and DQ settings. The underlying hypothesis is that SET may have special merits for certain subgroups of respondents. To report a specific sexual behavior can be more problematic for certain segments of the population, for example for women (Smith 1992) or religious people (see above); and hence SET could be particularly helpful for these people.[8]

Three regression models are estimated for each of the seven behaviors. An initial model is run for the whole sample including a dummy for question format (SET vs. DQ). This model yields the baseline effects of the covariates for the whole sample. Furthermore, this enables us to compare the bivariate treatment effect of SET (Table 3) with the one if other potential influence factors are held constant. Naturally, because of the experimental design, we would not expect any differences here; but because the experimental groups vary by some socio-demographic variables (Table 2), this might occur in our case. In a second step, separate models are run for the DQ and SET group. This will help to examine whether gender, age, partnership, etc. have different effects in the two question modes, i.e., whether there are interaction effects between question mode and the independent variables. The p–values of the corresponding significance tests for interactions are given in the last column of Tables 4, 5, 6, 7, 8, 9 and 10, which present the regression results. Tables 4, 5 and 6 and 10 report OLS-regressions, Tables 7, 8 and 9 binary logistic regressions. The two dependent variables "number of sexual partners in lifetime" and "coital frequency in the last four weeks" were transformed into logarithmic values because their distributions are heavily skewed.

Looking at the full models in Tables 4, 5, 6, 7, 8, 9 and 10, we find a confirmation of the bivariate result that SET makes a significant difference in response behavior in the case of age of first sex and masturbation. In SET mode, respondents report an older age of first sex and a higher masturbation frequency. No SET effect is found for the remaining sexual behaviors; this includes the number of lifetime sexual partners, for which no bivariate effect was found for the metric coding of the variable, but for the dummy coding only.

[8]Within research on social desirability, empirical studies have shown that often there is no consensus regarding which answer to a behavioral or attitudinal question is socially expected, i.e. corresponds to normative expectations (Stocké 2007a, b; Stocké and Hunkler 2007). Based on this finding of heterogeneous "SD beliefs," it is only a short step to the assumption that SET effects vary for different subgroups of the population.

Table 4 Determinants of the Number of Lifetime Sexual Partners (OLS Regressions)

	All	DQ	SET	P(Diff.)
SET (1 = yes)	0.02 (0.12)			
Female (1 = yes)	0.36* (2.18)	0.43 (1.61)	0.34 (1.63)	p = 0.78
Age (years)	0.12*** (5.40)	0.18*** (5.12)	0.08** (3.03)	p = 0.03*
Partnership (1 = yes)	0.63*** (3.81)	0.61* (2.26)	0.62** (2.99)	p = 0.96
Religiousness [1…5]	−0.29*** (4.17)	−0.45*** (3.80)	−0.21* (2.41)	p = 0.11
Party affinity [1…5]	0.52*** (5.92)	0.50*** (3.34)	0.52*** (4.84)	p = 0.92
Constant	−3.13*** (4.67)	−4.37*** (4.07)	−2.43** (2.84)	
Adj. R^2	0.13	0.20	0.10	
N	555	196	359	

Depicted are unstandardized regression coefficients and t-values in parentheses. The dependent variable has been transformed into logarithmic values.
DQ = direct questioning, SET = sealed envelope technique.
**: p < 0.05; **: p < 0.01; ***: p < 0.001*

The second row of Tables 4, 5, 6, 7, 8, 9, and 10 shows the gender effects. Being female is associated with reporting more sexual partners in one's lifetime than being male.[9] No significant gender differences can be observed for the age of first sex, coital frequency, one-night stands, and sexual infidelity. Significantly more women than men say that they have had homosexual experiences. Finally,

[9]This finding clearly contradicts other empirical studies having found that men report more sexual partners than women (Smith 1992; Tourangeau et al. 1997; Liljeros et al. 2001). We may speculate that our special sample of German university students could be responsible for this result. A widespread hypothesis about sex reporting is that men have a tendency towards overreporting whereas women have a tendency towards underreporting (Tourangeau et al. 1997, p. 355; Krumpal et al. 2018). However, in a recent study, Krumpal et al. (2018) do not find a clear pattern of gender-specific over- and underreporting of lifetime sexual partners, too.

Table 5 Determinants of Age of First Sex (OLS Regressions)

	All	DQ	SET	P(Diff.)
SET (1 = yes)	0.74** (3.20)			
Female (1 = yes)	−0.18 (0.79)	−0.36 (1.17)	−0.10 (0.33)	p = 0.57
Age (years)	0.12*** (4.15)	0.09* (2.14)	0.14*** (3.51)	p = 0.43
Partnership (1 = yes)	−0.99*** (4.40)	−0.84** (2.68)	−1.07*** (3.48)	p = 0.64
Religiousness [1…5]	0.40*** (4.11)	0.43** (3.17)	0.38** (2.93)	p = 0.80
Party affinity [1…5]	−0.53*** (4.37)	−0.34+ (1.92)	−0.60*** (3.77)	p = 0.32
Constant	15.12*** (16.40)	15.41*** (11.98)	15.66*** (12.61)	
Adj. R^2	0.14	0.12	0.12	
N	557	198	359	

Depicted are unstandardized regression coefficients and t-values in parentheses.
DQ = direct questioning, SET = sealed envelope technique.
*+: p < 0.1; *: p < 0.05; **: p < 0.01; ***: p < 0.001*

masturbation is much more common for men than for women, and this pattern is significantly better revealed in SET mode. In DQ mode, 27%, 48%, and 25% of male students answer that they never/seldom, sometimes, often/very often masturbate; these percentages change to 15%, 35%, and 50% in SET mode. The corresponding distributions for women amount to 43%, 50%, and 7% (via DQ), and 46%, 40%, and 14% (via SET). Going from DQ to SET, changes for men are bigger than those for women, and thus we can conclude that SET is particularly helpful for men to communicate about masturbation.[10]

[10]We also checked for interaction effects between the respondent's and interviewer's gender. We did not find any significant interaction effects with the exception of masturbation: If the interviewer is female and in SET mode only, the negative effect for female respondents is less pronounced as compared to a male interviewer.

Table 6 Determinants of Coital Frequency in the Last Four Weeks (OLS Regressions)

	All	DQ	SET	P(Diff.)
SET (1 = yes)	−0.04 (0.16)			
Female (1 = yes)	−0.25 (0.95)	−0.05 (0.12)	−0.39 (1.14)	p = 0.53
Age (years)	−0.01 (0.29)	−0.02 (0.42)	−0.01 (0.07)	p = 0.79
Partnership (1 = yes)	5.67*** (21.59)	6.22*** (15.59)	5.34*** (15.49)	p = 0.11
Religiousness [1…5]	−0.21+ (1.90)	−0.28 (1.62)	−0.17 (1.16)	p = 0.63
Party affinity [1…5]	0.41** (2.94)	0.72** (3.20)	0.27 (1.49)	p = 0.13
Constant	−4.62*** (4.30)	−5.35*** (3.33)	−4.31** (3.04)	
Adj. R^2	0.46	0.57	0.40	
N	557	200	357	

Depicted are unstandardized regression coefficients and t-values in parentheses. The dependent variable has been transformed into logarithmic values.
DQ = direct questioning, SET = sealed envelope technique.
*+: p < 0.1; **: p < 0.01; ***: p < 0.001*

The third row of Tables 4, 5, 6, 7, 8, 9 and 10 informs us about the age effects. According to Table 4, older students report more lifetime sexual partners than younger ones, which is no surprising result. This pattern is clearly more pronounced in DQ than in SET mode. Hence, in the direct face-to-face situation older students seem to overreport the diversity of their sexual life. This interpretation is supported by looking at the probability of one-night stands. Again, older students report more of them, but only in the DQ version; the difference vis-à-vis SET is significant with p = 0.05. There is also a significant positive age effect with respect to the age of first sex. For coital frequency, sexual infidelity, homosexual contacts and masturbation, age effects are insignificant.

Regarding partnership, i.e. whether the respondent is currently living in a partnership or not, several results are interesting in Tables 4, 5, 6, 7, 8, 9 and 10. Those currently in a partnership can look back to more sexual partners in their life to date. This is of course just a descriptive result; the causal relationship is

Table 7 Determinants of One-Night Stand (Binary Logistic Regressions)

	All	DQ	SET	P(Diff.)
SET (1 = yes)	−0.16 (0.89)			
Female (1 = yes)	0.12 (0.66)	0.24 (0.79)	0.05 (0.24)	p = 0.62
Age (years)	0.07** (2.92)	0.15** (3.02)	0.04 (1.25)	p = 0.05+
Partnership (1 = yes)	−0.14 (0.81)	−0.58+ (1.88)	0.09 (0.40)	p = 0.08+
Religiousness [1…5]	−0.26*** (3.36)	−0.37** (2.76)	−0.2*1 (2.21)	p = 0.33
Party affinity [1…5]	0.37*** (3.81)	0.42* (2.46)	0.33** (2.84)	p = 0.67
Constant	−1.95** (2.63)	−3.48* (2.52)	−1.37 (1.51)	
Adj. R^2	0.04	0.09	0.03	
N	568	207	361	

Depicted are unstandardized logit coefficients and z-values in parentheses.
DQ = direct questioning, SET = sealed envelope technique.
*+: p < 0.1; *: p < 0.05; **: p < 0.01; ***: p < 0.001*

probably the other way round, meaning that past sexual experience (which in turn may be influenced by variables like attractivity) positively affects the probability of living in a partnership today. Not surprisingly, they also have a higher coital frequency in the last four weeks. The partnership effect in Table 6 is overwhelmingly strong, and, based on this effect, the explained variance of these models is much better than in the other models. Although the difference is not significant (p = 0.11), the regression coefficient of partnership is higher in DQ than in SET format. It can be assumed that it is socially expected that partners have regular intercourse, and this norm probably triggers an overreporting of coital frequency by respondents with a regular life partner. Hence, the lower SET coefficient is probably a better estimate of the true value. There are also significant effects of a partnership for the age of first sex; respondents living in a regular partnership report having had their first sex being roughly one year younger than singles. For sexual infidelity, the partnership effects are not significant. Regarding one-night stands, homosexual contacts, and masturbation, there are significant differences

Table 8 Determinants of Sexual Infidelity in Current/Last Partnership (Binary Logistic Regressions)

	All	DQ	SET	P(Diff.)
SET (1 = yes)	−0.16 (0.56)			
Female (1 = yes)	−0.31 (1.13)	−0.27 (0.62)	−0.36 (0.99)	p = 0.88
Age (years)	0.01 (0.03)	−0.05 (0.77)	0.03 (0.65)	p = 0.32
Partnership (1 = yes)	−0.27 (0.97)	−0.35 (0.80)	−0.24 (0.66)	p = 0.84
Religiousness [1…5]	−0.06 (0.47)	−0.10 (0.47)	−0.04 (0.27)	p = 0.84
Party affinity [1…5]	0.18 (1.22)	0.22 (0.91)	0.17 (0.88)	p = 0.86
Constant	−2.01+ (1.74)	−0.71 (0.36)	−2.88* (1.97)	
Adj. R^2	0.01	0.02	0.01	
N	540	199	341	

Depicted are unstandardized logit coefficients and z-values in parentheses.
DQ = direct questioning, SET = sealed envelope technique.
*+: p < 0.1; *: p < 0.05*

between DQ and SET (although at 10% levels only). Here, the response pattern is such that the partnership effects in DQ mode reflect an underreporting of one-night stands, homosexual contact, and masturbation; all of which being supposedly socially undesirable for respondents living in a regular partnership (even if the behavior pertains to something that has happened before the actual partnership). We can draw the conclusion that SET is particularly helpful for those living in a partnership to give more valid answers to sensitive questions pertaining to sexual behavior.

Similarly, SET makes it easier for religious people to communicate honestly about their own sexual behavior. This can be seen in the fifth row of Tables 4, 5, 6, 7, 8, 9 and 10. For each behavioral item, the DQ effects of religiosity are negative (except age of first sex, where it is, as suspected, positive) and, hence, are in line with the supposed direction of social desirability for religious people (which consists in not or not too exceedingly perform the respective sexual behaviors). The SET effects do not change from minus to plus, but all SET effects

Table 9 Determinants of Homosexual Contact (Binary Logistic Regressions)

	All	DQ	SET	P(Diff.)
SET (1 = yes)	0.27 (1.08)			
Female (1 = yes)	1.08*** (4.05)	1.39** (2.83)	0.97** (3.03)	p = 0.47
Age (years)	−0.02 (0.65)	−0.07 (0.93)	−0.01 (0.39)	p = 0.52
Partnership (1 = yes)	0.62* (2.49)	−0.06 (0.15)	0.97** (2.97)	p = 0.05+
Religiousness [1…5]	−0.15 (1.49)	−0.49* (2.40)	−0.04 (0.30)	p = 0.06+
Party affinity [1…5]	−0.09 (0.74)	−0.08 (0.34)	−0.11 (0.71)	p = 0.91
Constant	−1.75+ (1.74)	0.16 (0.08)	−2.07+ (1.76)	
Adj. R^2	0.06	0.09	0.07	
N	572	209	363	

Depicted are unstandardized logit coefficients and z-values in parentheses.
DQ = direct questioning, SET = sealed envelope technique.
*+: p < 0.1; *: p < 0.05; **: p < 0.01; ***: p < 0.001*

are lower in size than the DQ effects, and significantly so for homosexual contact and masturbation (on a 10% level). Thus, although the reduction in SET is not significant for the rest of the items, the consistency of the change pattern can be qualified as a finding in favor of SET. This is furthermore confirmed when a global test of the religiousness effects across all items is conducted. Similar to the procedure of the global test in Table 3, metric variables were coded into dummies (the dummy for "age of first sex" now indicating a young age, less than 18 years) and the data transformed into long format. Two logistic regression models using standard errors adjusted for clustering by respondent yield an overall DQ effect of religiousness of −0.15 (p = 0.001), and an overall SET effect of religiousness of −0.02 (p = 0.579, n.s.). The difference between the effects (difference-in-differences) is significant with z = 2.22 and p = 0.026.

Focusing on students who like to go to university parties, the sensitivity assumption of the sex questions may be more problematic than for those who do not go to parties. It could be argued that many party girls and boys hold values that are in favor of a liberal sexual lifestyle. Such a more liberal lifestyle shows up in Tables 4, 5, 6, 7, 8, 9 and 10 in the form that "party persons" report more

Table 10 Determinants of Masturbation Frequency (OLS Regressions)

	All	DQ	SET	P(Diff.)
SET (1 = yes)	0.34*** (4.06)			
Female (1 = yes)	−0.75*** (9.22)	−0.45*** (3.59)	−0.91*** (8.74)	p = 0.01**
Age (years)	0.01 (0.63)	0.02 (1.44)	−0.01 (0.35)	p = 0.19
Partnership (1 = yes)	−0.16+ (1.95)	−0.34** (2.71)	−0.06 (0.61)	p = 0.10+
Religiousness [1…5]	−0.11** (3.16)	−0.20*** (3.66)	−0.07 (1.56)	p = 0.07+
Party affinity [1…5]	0.05 (1.21)	0.10 (1.40)	0.02 (0.36)	p = 0.39
Constant	3.11*** (9.34)	2.70*** (5.27)	3.77*** (8.79)	
Adj. R^2	0.18	0.15	0.18	
N	553	198	355	

Depicted are unstandardized regression coefficients and t-values in parentheses. The dependent variable "masturbation frequency" is measured on a 5-digit scale from 1 = never to 5 = very often.
DQ = direct questioning, SET = sealed envelope technique.
*+: p < 0.1; **: p < 0.01; ***: p < 0.001*

sexual partners in their lifetime, a younger age of first sex, a higher intercourse frequency in the last four weeks, and more one-night stands. Comparing DQ and SET effects, the findings in Table 6 suggest that in the DQ setting party persons tend to declare more sex than they really have, but this speculation does not find further support in the light of the other behavioral items.

4 Results 2: Additional Findings

Besides the main evaluation criterion whether SET contributes to a reduction of the supposed underreporting bias of sensitive behavior in general and for special subgroups, this section will present results on three additional evaluation criteria: 1) the level of item nonresponse, 2) the level of respondent's uneasiness, and 3) the level of interviewer's uneasiness.

A widespread assumption with respect to sensitive questions is that respondents refuse to answer such questions (e.g., Lensvelt-Mulders 2008). This results in missing values, and because these missings are not "at random", estimates of population prevalences are confronted with an additional difficulty. Previous research has shown, however, first, that for many sensitive topics, item nonresponse is surprisingly low (Wolter and Herold 2018; Preisendörfer and Wolter 2014; Wolter and Laier 2014), and, second, that the level of item nonresponse cannot be taken as a good indicator of the degree of sensitivity of a topic. Item nonresponse to the income question, for instance, is usually much higher than nonresponse to sex questions like the number of sexual partners in a person's lifetime or coital frequency in the last four weeks (Tourangeau et al. 2000, pp. 260–264). We will nevertheless explore whether item-nonresponse presents a problem regarding the sex questions, and, in case it does, whether SET is helpful in reducing it. Table 11 shows the results for the seven sex items under investigation.

In line with prior research, item nonresponse is generally low. The highest value is about 7% for the number of sexual partners in DQ mode. Based on their level of nonresponse, the seven items can be grouped into two clusters: number of sexual partners, age of first sex, coital frequency, and masturbation with above average nonresponse rates, and one-night stands, sexual infidelity, and

Table 11 Item-Nonresponse by Question Mode

	DQ (n = 211)	SET (n = 367)	Diff: χ^2 or t
Sexual partners (lifetime)	6.6	1.6	10.03**
Age of first sex	5.7	1.6	7.29**
Coital frequency (last 4 weeks)	4.7	2.2	2.91^{+}
One-night stand (ever)	1.4	1.1	0.12
Sexual infidelity in current/last relationship (ever)	0.5	0.0	1.70
Homosexual contact (ever)	0.5	0.8	0,23
Masturbation	5.7	2.7	3.21^{+}
Average number of missing values [0…7]	0.25	0.10	2.32*

Depicted are percent values and mean values in the last row. Differences by question mode were tested for by using χ^2-tests and t-tests (last row).
DQ = direct questioning, SET = sealed envelope technique.
*+: $p < 0.1$; *: $p < 0.05$; **: $p < 0.01$*

Table 12 Respondents' Subjectively Perceived Uneasiness when Answering Sensitive Sex Questions

Level of uneasiness	DQ (n = 211)	SET (n = 367)
Low	52.6	67.6
Moderate	37.4	28.0
High	10.0	4.4
Diff: χ^2	15.0***	

Depicted are percent values.
DQ = direct questioning, SET = sealed envelope technique.
***: $p < 0.001$*

homosexual contacts with below average rates.[11] The expectation that SET is associated with lower refusal rates is also confirmed. The DQ-SET differences are significant at least on a ten percent level for four of the items. The average number of missing values over the six items (empirical range 0–6) is 0.25 for DQ and 0.1 for SET, and this again is a significant difference. Hence, the overall result regarding nonresponse is again in favor of SET, which is in line with the findings of the previous section.

A further evaluation criterion of the virtues of SET are feelings of uneasiness on the part of the respondents when being asked about intimate topics like sexual behaviors. According to the theoretical remarks in Sect. 1, SET can be seen as a device that aims at reducing the level of respondents' uneasiness by allowing more privacy. At the end of our questionnaire, we attempted to ascertain whether SET was actually successful in reaching this aim. Using the above (Table 1) presented indicator of respondents' subjectively perceived degree of uneasiness when answering the sensitive questions, Table 12 presents the recoded frequency distributions for DQ and SET. The level of uneasiness is not exceedingly high in both question modes. Nevertheless, the difference between DQ and SET is highly significant in the predicted direction. The same holds if we do not recode the 10-digit scale and compare the mean values which amount to 3.8 in DQ and to

[11]Note that the questions concerning one-night stands, sexual infidelity and homosexual contacts are formulated with the frame "ever engaged in." Such a frame is normally less threatening than the frame "currently engaged in." Moreover, these questions are simple yes/no questions as opposed to the other questions which ask—in a more demanding manner—for numbers and frequencies.

2.9 in SET mode, respectively ($t = 4.79$, $p < 0.001$). Therefore, in line with expectations, SET reduces the level of interviewee's feelings of discomfort.

With regard to the last criterion "level of interviewer's uneasiness," only selective qualitative findings can be described. These consist in impressions gained in group discussions within the seminar in which the study was carried out. Based on the finding of prior research that sensitive questions are sometimes more a problem for interviewers than for interviewees (see Sect. 1), the expectation was that interviewers prefer the SET to the DQ procedure. This expectation was already supported in the process of assigning interviewers to the two modes. Assignment was at random (see Sect. 2), and, given the results of the assignment procedure, none of the interviewers in the SET group but several interviewers in the DQ group began to articulate objections and suggested allowing voluntary group changes. This means that DQ interviewers anticipated more difficulties and inconveniences than SET interviewers. And indeed, these anticipations did find justification. In seminar discussions after the field work, more DQ than SET interviewers said that they did not feel good when it came to discussing sex, that they had to give additional explanations, excuses, allow rough estimates, and so on. Nevertheless, as mentioned already in Sect. 2, unit nonresponse did not significantly differ between DQ and SET.

5 Conclusion

On the basis of the findings presented in this study, we can conclude that SET seems to have some merits in face-to-face interviews as compared to conventional direct questions. It is helpful in alleviating misreporting, especially for highly sensitive topics (like masturbation); it is a special help for certain subgroups of respondents (like religious people or people living in a partnership); it reduces item-nonresponse; and it lowers the level of uneasiness on the side of the interviewer and the respondents.

Of course, this study has limitations. The three most evident are: First, it is confined to the area of sexual behavior and, thus, leaves out sensitive topics in other areas. We also did not elaborate how sexual issues relate to the more general field of deviant behavior. Second, our sample is restricted to university students in one single German city; hence, inference to any type of general population is not possible. Given the fact that most parts of our questionnaire touched sensitive topics ("sensitive topic survey"), it was surprising how unproblematic the field work of the survey was. It can be assumed that ordinary people will see sex questions (such as those used in our survey) more skeptically than students. This means

that the average level of sensitivity of the same questions would be higher, and the prediction would be that SET would be even more advantageous than in our survey. Whether this prediction holds remains to be investigated in future work. Third, we based our argumentation on the "more-is-better" (or "less-is-better", depending on the direction of social desirability) assumption. This is because we are lacking true values of the sexual behaviors investigated in our study—which could have been provided by external individual validation data. Such data, however, are usually very difficult to acquire. Hence, even if our SET estimates point into the assumed direction as compared to their DQ counterparts, they might still miss the true values. All in all, the results of this study should be taken as evidence in favor of SET without claiming that SET is the all-time perfect and trouble-free device solving all problems related to sensitive questions.

One general objection against SET could be that its merits simply reflect the advantages of self-administered mail or web surveys (greater belief in anonymity and confidentiality, no interviewer effects).[12] This is partly true, but to switch completely to a self-administered mail or web surveys does not necessarily seem to be good advice for sensitive topics. A "sensitive topic survey" (like ours) via mail or web would probably have a low response rate, i.e. unit nonresponse would be high. Furthermore, via mail or web, control over the response situation would be lower (who answered the questions, how serious the answers are). SET makes sense for face-to-face surveys only. For this type of survey, it incorporates the merits of self-administration without losing the advantages of face-to-face interviews.

References

Aquilino, W. S. (1998). Effects of interview mode on measuring depression in younger adults. *Journal of Official Statistics, 14*(1), 15–29.

Auspurg, K., & Hinz, T. (2014). *Factorial survey experiments (quantitative applications in the social sciences 175)*. Thousand Oaks: Sage.

Barnett, J. (1998). Sensitive questions and response effects: An evaluation. *Journal of Managerial Psychology, 13*(1–2), 63–76.

[12]With respect to sex questions, Tourangeau et al. (1997), for example, have demonstrated that self-administered questions (SAQs) produce higher levels of reporting. See, for example, Tourangeau and Yan (2007); Wolter (2012, pp. 57–61) for literature reviews regarding the effects of self-administration (and other survey modes) on response behavior.

Barton, A. H. (1958). Asking the embarrassing question. *Public Opinion Quarterly, 22*(1), 67–68.

Becker, R. (2006). Selective responses to questions on delinquency. *Quality & Quantity, 40*(4), 483–498.

Becker, R., & Günther, R. (2004). Selektives Antwortverhalten bei Fragen zum delinquenten Handeln. Eine empirische Studie über die Wirksamkeit der „sealed envelope technique" bei selbst berichteter Delinquenz mit Daten des ALLBUS 2000. *ZUMA-Nachrichten, 54,* 39–59.

Benson, L. E. (1941). Studies in secret-ballot technique. *Public Opinion Quarterly, 5*(1), 79–82.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken: Wiley-Interscience.

Bradburn, N. M., & Sudman, S. (1979). *Improving interview method and questionnaire design. Response effects to threatening questions in survey research*. San Francisco: Jossey-Bass.

Cantril, H. (1940). Experiments in the wording of questions. *Public Opinion Quarterly, 4*(2), 330–332.

De Leeuw, E. D., & Edith, D. (2001). Reducing missing data in surveys: An overview of methods. *Quality & Quantity, 35*(2), 147–160.

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 (Hrsg.), *Measurement errors in surveys* (S. 185–210). New York: Wiley.

Esser, H. (1991). Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als „rational choice". In R. Wittenberg (Hrsg.), *Person—Situation—Institution—Kultur. Günter Büschges zum 65. Geburtstag* (S. 59–78). Berlin: Duncker & Humblot.

Fox, J. A., & Tracy, P. E. (1986). *Randomized response. A method for sensitive surveys*: Vol. 07–058. *Sage university paper series on quantitative applications in the social sciences*. Newbury Park: Sage.

Hox, J. J., & Leeuw, E. D. D. (2002). The influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Hrsg.), *Survey nonresponse* (S. 103–120). New York: Wiley.

Jann, B., I. Krumpal, & F. Wolter (Ed.). (2019). *Social desirability bias in surveys—collecting and analyzing sensitive data*: vol. 13. mda—Methods, Data, Analyses.

Jasso, G. (1985). Marital coital frequency and the passage of time: Estimating the separate effects of spouses' ages and marital duration, birth and marriage cohorts, and period influences. *American Sociological Review, 50*(2), 224–241.

Krosnick, J. A., Holbrook, A. L., Berendt, M. K., Carson, R. T., Michael Hanemann, W., Kopp, R. J., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Moody, W. R., Green, M. C., & Conaway, M. (2002). The impact of „no opinion" response options on data quality. Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly, 66*(3), 371–403.

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

Krumpal, I., Jann, B., Korndörfer, M., & Schmukle, S. C. (2018). Item sum double-list technique: An enhanced design for asking quantitative sensitive questions. *Survey Research Methods, 12*(2), 91–102.

Lee, R. M. (1993). *Doing research on sensitive topics*. Thousand Oaks: Sage.

Lensvelt-Mulders, G. J. L. M. (2008). Surveying sensitive topics. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Hrsg.), *International handbook of survey methodology* (S. 461–478). New York: Lawrence Erlbaum.

Liljeros, F., Edling, C. R., Nunes, L. A., Eugene Amaral, H., & Aberg, Y. (2001). The web of human sexual contacts. *Nature, 411*, 907–908.

Makkai, T., & McAllister, I. (1992). Measuring social indicators in opinion surveys: A method to improve accuracy on sensitive questions. *Social Indicators Research, 27*(2), 169–186.

Perry, P. (1979). Certain problems in election survey methodology. *Public Opinion Quarterly, 43*(3), 312–325.

Preisendörfer, P. (2008). *Heikle Fragen in mündlichen Interviews: Ergebnisse einer Methodenstudie im studentischen Milieu* (unpublished manuscript). Mainz: Johannes Gutenberg-Universität Mainz.

Preisendörfer, P., & Wolter, F. (2014). Who is telling the truth? A validation study on determinants of response behavior in surveys. *Public Opinion Quarterly, 78*(1), 126–146.

Rasinski, K. A., Willis, G. B., Baldwin, A. K., Yeh, W., & Lee, L. (1999). Methods of data collection, perceptions of risks and losses, and motivation to give truthful answers to sensitive survey questions. *Applied Cognitive Psychology, 13*(5), 465–484.

Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen*. Opladen: Leske + Budrich.

Schnell, R., & Kreuter, F. (2005). Separating interviewer and sampling-point effects. *Journal of Official Statistics, 21*(3), 389–410.

Smith, T. W. (1992). Discrepancies between men and women in reporting number of sexual partners: A summary from four countries. *Social Biology, 39*(1–2), 203–211.

Stocké, V. (2004). Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie, 33*(4), 303–320.

Stocké, V. (2007a). Determinants and consequences of survey respondents' social desirability beliefs about racial attitudes. *Methodology, 3*(3), 125–138.

Stocké, V. (2007b). The interdependence of determinants for the strength and direction of social desirability bias in racial attitude surveys. *Journal of Official Statistics, 23*(4), 493–514.

Stocké, V., & Hunkler, C. (2007). Measures of desirability beliefs and their validity as indicators for socially desirable responding. *Field Methods, 19*(3), 313–336.

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine.

Sudman, S., & Bradburn, N. M. (1982). *Asking questions. A practical guide to questionnaire design*. San Francisco: Jossey-Bass.

Tourangeau, R., Rasinski, K., Jobe, J. B., Smith, T. W., & Pratt, W. F. (1997). Sources of error in a survey on sexual behavior. *Journal of Official Statistics, 13*(4), 341–365.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions. The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60*(2), 275–304.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859–883.

Wolter, F. (2012). *Heikle Fragen in Interviews. Eine Validierung der Randomized Response-Technik*. Wiesbaden: Springer VS.

Wolter, F., & Herold, L. (2018). Testing the item sum technique (ist) to tackle social desirability bias. *SAGE Research Methods Cases*. doi:http://dx.doi.org/10.4135/9781526441928.

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*(3), 153–168.

Wolter, Felix Dr., Institut für Soziologie, Johannes Gutenberg-Universität Mainz, Jakob-Welder-Weg 12, D-55128 Mainz, felix.wolter@uni-mainz.de. Forschungsschwerpunkte: Methoden der empirischen Sozialforschung, Sozialstruktur und soziale Ungleichheit, Kompetenzdiagnostik, Rational-Choice- und Spieltheorie.

Preisendörfer, Peter Prof. Dr., Institut für Soziologie, Johannes Gutenberg-Universität Mainz, Jakob-Welder-Weg 12, D-55128 Mainz, preisendoerfer@uni-mainz.de, Forschungsschwerpunkte: Organisations- und Arbeitsmarktforschung, Sozialwissenschaftliche Umweltforschung, Methoden der empirischen Sozialforschung.