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

Peter Preisendörfer
Institute of Sociology
University of Mainz
Germany
E-Mail: preisendoerfer@uni-mainz.de

October 2008

---

* This article is part of the author's research within the research project "Asking Sensitive Questions" supported by the German Research Foundation DFG (Priority Program SPP 1292, Grant PR 237/6-1). For helpful comments I would like to thank Claudia Diehl and Andreas Diekmann.

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

Abstract: A "sensitive topic survey" of students (n=578) of the University of Mainz in Germany was conducted to evaluate the merits of the sealed envelope technique in face-to-face-interviews. 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 underreporting of highly 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 interviewees 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.

Keywords: Sensitive questions, sealed envelope technique, social desirability, misreporting in surveys, measurement error, sexual behavior.

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

1. Introduction

Sensitive questions, i.e. questions that pertain to personal, secret, socially frowned upon or illegal behavior, constitute a special challenge for the theory and practice of survey research. This is because many problems of conventional surveys (like item nonresponse or social desirability) are revealed in research on sensitive topics in a more pronounced and more severe form. Obtaining valid answers to sensitive 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. Raymond Lee (1993: 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 (for reviews, Lee 1993; Barnett 1998; Tourangeau et al. 2000: Chap. 9; Tourangeau and Yan 2007), 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 thus 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, normally in the form of underreporting? 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, sealed envelopes and vignettes to the item count and randomized response technique. 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, detailed methodological reports about experiences with this technique are rare (see, however, Makkai and Mcallister 1992; Becker and Günther 2004, 2006). 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: 281-284; Stocké 2007a), we can expect interviewees to decide to answer 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: 3-11; Tourangeau et al. 2000: 257-259). Respondents are interested in making a good impression on the interviewer and 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 interviewee (Bradburn and Sudman 1979: Chap. 4; Schnell 1997: 274-275; Hox and de Leeuw 2002; Biemer and Lyberg 2003: 180; Schnell and Kreuter 2005). The baseline assumption behind SET is that its application makes it easier for the interviewer and

[1] With respect to the social desirability bias, rational choice theory focuses on three determinants of this misreporting bias (Stocké 2007a): (1) respondents' 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.

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 in general and for certain subgroups of respondents a question is, the more useful in comparison to direct questioning SET will be.

To examine this hypothesis, this article refers to sexual behavior, a classic field of sensitive topic research (Tourangeau et al. 1997). Based on a survey with 578 students of the University of Mainz (Germany), SET answers will be compared to DQT answers, i.e. answers given in response to direct questions (DQT=direct questioning technique). The main criterion for the comparison and for the evaluation of SET is the presumed underreporting bias. To what extent is SET successful in overcoming the underreporting associated with direct questions? The underreporting problem will be dealt with in simple descriptive as well as multivariate analyses. Besides the underreporting criterion, some additional and supplementary criteria will be used to evaluate the merits of SET.

In the next section, necessary information about the empirical data will be given. Section 3 presents the descriptive findings of the prevalence of different sexual behaviors – the SET prevalences as compared to the DQT prevalences. Even if both prevalences are identical, it may be that factors influencing them are different for the two question modes. This will be explored via multivariate regression analyses in Section 4. Additional criteria relevant for the assessment of SET are discussed in Section 5. The article ends with a short conclusion.

2. Empirical Data

The data used in this article come from a small-scale research project initiated by the author in collaboration with his students in a one-year seminar on social research methods. In October and November 2007, a group of 578 students of the University of Mainz (Germany) were interviewed by the 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 are registered at the University of Mainz. 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, postgraduate 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%. Undoubtedly, this cannot be qualified as a good result.

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 was given). 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,

[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.

including sensitive ones.

There were two versions of the questionnaire: In the first one, the "direct version," all sensitive questions were asked directly. The second one, the "indirect version", applied different techniques to ask the sensitive topics in a more or less elaborated way. Besides wording/framing techniques (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] In line with a split-ballot design, the direct version was intended to give the baseline and the indirect version the comparison values. The overall expectation was that the indirect SET version would result in higher prevalence values than the DQT 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, 367 (63%) indirect ones.[4] Concerning response rates (and thus 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."

[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.

3. Descriptive Findings

The "sex module" of our questionnaire was located near the end of the interview (prior to the closing socio-demographic questions). It covered 20 questions, but only six of them can be qualified as core questions. The other questions served as an introduction, were tailored to restricted subgroups, or had other special purposes in the interview process. In the direct version and the indirect SET version, the questions were identical in layout and wording.

The interviewers of the SET application had an envelope marked "Confidential." Inside this envelope was a four-page questionnaire with the 20 sex questions. 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." Comparing the direct question technique (DQT) and the sealed envelope technique (SET), Table 1 presents the descriptive findings of the six questions of interest, connected with tests for significance of the differences.[5]

Table 1 about here

Concerning the number of sexual partners in their lifetime to date, no significant difference can be observed between DQT and SET. Nevertheless, an inspection of the frequency distributions 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. (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. 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 DQT and SET distributions: Recoding the number of sexual partners as low (0-2 partners), moderate (3-6 partners) and high (7 or more partners) yields a DQT distribution of 31%, 43% and 26% respectively and a SET distribution of 35%, 34%

[5] For the first two items in Table 1 (no. of sexual partners, coital frequency), the significance tests are t-tests. For the other items, chi-squared tests are used.

and 31% respectively. 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.

Also no significant differences between DQT and SET can be found in Table 1 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, 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.[6] 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, DQT and SET yield percentages of masturbation which clearly differ. In the direct question mode, 15% admit to masturbating often or very often, compared to 29% in SET mode, i.e. the percentage value almost 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" compiled by Bradburn and Sudman (1979: 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 1 (from sexual partners to homosexual contacts) it could be argued that their sensitivity is low, nonlinear (as for the number of sexual partners) or even that these are not at all sensitive. Masturbation, however, is undoubtedly a sensitive issue.

[6] 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).

4. Multivariate Findings

We saw from Table 1 that – with the (significant) exception of masturbation – SET did not result into higher prevalence values of sexual behaviors than DQT. Even if prevalence values are similar in bivariate comparisons, it is still possible, however, that SET influences the response pattern after controlling for other determinants of the behavior of interest. Furthermore, these other determinants may be of different strength in the SET and the DQT settings. This means that SET may have special merits for certain subgroups of respondents. To report a behavior can be more problematic for certain segments of the population, for example, for women or religious people, and hence SET could be particularly helpful for these people.[7] In technical terms, we can investigate this by estimating multiple regression models with the behaviors in Table 1 as dependent variables and a set of proposed independent variables.

Concerning the selection of independent variables, we chose – besides the dummy 1=SET and 0=DQT – five covariates: (1) gender, coded as dummy with 1=female and 0=male: 58% of the sample are female, 42% male; (2) age, measured in years: although most of the students interviewed are in their twenties, there is an age range from 19 to 42, the mean age is 25; (3) partnership, coded as dummy with 1=currently living in a partnership and 0=currently not living in a partnership: 58% of the respondents are currently in a partnership; (4) religious affiliation, based on a question whether the respondents declared themselves to be religious on a 5-digit scale from 1=not religious at all to 5=very strong religious affiliation: 58% answer that they have no or a low religious affiliation, 23% moderate, and 19% strong or very strong; (5) partying habits, based on a question how often respondents go to student parties in the evenings, measured on a 5-digit scale from 1=never to 5=very often: 48% say that they never or seldom go to such parties, 39% sometimes, and 13% often or very often. Gender, age and partnership have been selected as independent variables for the regression models 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. And partying habits are seen as a mixed indicator of personal needs as well as situational opportunities for sexual experiences.

[7] Within research on social desirability, empirical studies have shown that (surprisingly) often there is no consensus on which answer to a behavioral or attitudinal question is socially expected, i.e. corresponds to normative expectations (e.g., Stocké 2007a, 2007b; 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.

Three regression models are estimated for each of the six behaviors. An initial model (Model 1) includes the full set of the covariates. This enables us to see whether SET affects the response behavior while keeping other potential influence factors constant. In a second step, separate models are run for the DQT and the SET groups (Models 2a and 2b). This will help to examine whether gender, age, partnership, etc. have similar effects in the two question modes. To estimate separate regression models for the DQT and the SET groups means that we are looking for interaction effects. If the interaction effects are significant, corresponding regression coefficients for the two modes are different. The p-values of these significance tests (t-tests) are given in the last column of Tables 2-7 which present the regression results. Tables 2, 3 and 7 report OLS-regressions, Tables 4, 5 and 6 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.

Tables 2-7 about here

Looking at the first row of the six tables, we find a confirmation of the bivariate result that SET makes a significant difference in response behavior only in the case of masturbation. The SET effect in Model 1 of Table 7 is positive and highly significant.

The second row of Tables 2-7 shows the gender effects. Being female is associated with reporting more sexual partners in one's lifetime than being male.[8] No gender differences can be observed for coital frequency, one-night stands and sexual infidelity. Significantly more women than men say that they have had homosexual experiences. Finally, masturbation is much more common for men than for women, and this pattern is significantly better revealed in SET mode. In DQT mode, 27%, 48% and 25% respectively of the men answer that they never/seldom, sometimes, often/very often masturbate; these percentages change to 15%, 35% and 50% respectively in SET mode. The corresponding distributions for the women are 43%, 50% and 7% (via DQT) and 46%, 40% and 14% (via SET). Going from DQT to SET, changes for men are greater than those for women, and thus we can conclude that SET is particularly helpful for men to communicate about masturbation.

[8] This finding clearly contradicts other empirical studies which have found that men report more sexual partners (Smith 1992; Tourangeau et al. 1997). 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: 355).

The third row of Tables 2-7 informs us about the age effects. According to Table 2, older students report more sexual partners in their lifetime than younger ones. This pattern is clearly more pronounced in DQT 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, but only and especially in the direct question version. For coital frequency, sexual infidelity, homosexual contacts and masturbation, age effects are insignificant.

With respect to partnership, i.e. whether the respondent is currently living in a partnership or not, several results are interesting in Tables 2-7. Those currently in a partnership can look back to more sexual partners in their life to date. Not surprising, they also have a higher coital frequency in the last four weeks. The partnership effect in Table 3 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 the DQT than in the SET format. It is socially expected that partners have regular intercourse, and this expectation is more clearly reflected in the DQT responses. Based on common knowledge, however, it can be argued that the lower SET coefficient is probably a better estimate of the "real value." For one-night stands and sexual infidelity, the partnership effects are not significant, but again the DQT responses are more in line with social desirability. The same holds true for homosexual contacts. Whereas in the light of DQT respondents with and without a partner do not differ in their homosexual experiences, there is a significant difference in the light of SET. And once more for masturbation: According to DQT, respondents with a partner masturbate less often, but this is not so in SET mode. 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 sixth row of Tables 2-7. For each behavioral item, the DQT effects of religiosity are negative and, hence, are in line with social desirability (defined as follows: you should have a low number of sexual partners, a low coital frequency, no one-night stands, no sexual infidelity in partnerships, no homosexual contacts, a low masturbation frequency). The SET effects do not change from minus to plus, but all SET effects are lower in size than the DQT effects. Although in none of the six tables the reduction is significant, the consistency of the change pattern can be qualified as a remarkable finding.

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 2-7 in the form that “party persons” report more sexual partners in their lifetime, a higher intercourse frequency in the last four weeks and more one-night stands. Comparing DQT and SET effects, the findings in Table 3 suggest that in the DQT 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.

5. Additional Findings

Besides the main evaluation criterion whether SET contributes to a reduction of the supposed underreporting bias of sensitive behavior in general (Section 3) and for special subgroups (Section 4), this section introduces three additional evaluation criteria: (1) the level of item nonresponse, (2) the level of interviewer's uneasiness, and (3) the level of respondent's uneasiness.

A widespread assumption with respect to sensitive questions is that respondents refuse to answer such questions. The result is 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, and, second, that the level of item nonresponse cannot be taken as a good indicator of the degree of sensitivity of a topic (Tourangeau et al. 2000: 260-264). Item nonresponse to the income question, for example, 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: 263-264). Nevertheless, there is a positive correlation between sensitivity of a topic and the corresponding level of item nonresponse. Based on this finding, SET would have an advantage to DQT if it yields fewer missing values. Table 8 shows the results for the six sex items under investigation.[9]

Table 8 about here

In line with prior research, item nonresponse is generally low. The highest value is 7% for the number of sexual partners in DQT mode. Based on their level of nonresponse, the six items can be grouped into two clusters: number of sexual partners, coital frequency and masturbation with above average refusal rates, and then one-night stands, sexual infidelity and homosexual contacts with below average refusal rates. This grouping roughly matches general knowledge about the degree of sensitivity of these topics for German university students, and hence confirms the prediction of a positive association between sensitivity and item nonresponse.[10]

[9] The last column of Table 8 is based on chi-squared tests to examine significance of the differences. An exception is the last row which shows the result of a t-test.

[10] 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.

The expectation, of primary interest, that SET is associated with lower refusal rates is also confirmed. The DQT/SET differences are significant or nearly significant for the three items with "high sensitivity" (as measured by their general level of item nonresponse), whereas there are no systematic differences for the "low sensitivity" items. The average number of missing values over the six items (range 0-6) is 0.19 for DQT and 0.08 for SET, and this again is a significant difference. This finding is in line with the more general proposition: "When the topic is sensitive or seen as threatening to the respondent, the use of self-administered questionnaires reduces the number of missing data" (De Leuuw 2001: 152).

With regard to the criterion "level of interviewer's uneasiness," only qualitative impressions can be described – impressions gained in group discussions within the seminar responsible for the study.[11] Based on the finding of prior research that sensitive questions are sometimes more a problem for interviewers than for interviewees (see Section 1), the expectation was that interviewers prefer the SET to the DQT mode. This expectation was already supported in the process of assigning interviewers to the two modes. Assignment was at random (see Section 2), and, given the results of the assignment procedure, none of the interviewers in the SET group but several interviewers in the DQT group began to articulate objections and suggested allowing voluntary group changes. This means that DQT interviewers anticipated more difficulties and inconveniences than SET interviewers. And indeed, these anticipations did find justification. In seminar discussions after the field work, more DQT 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 Section 2, unit nonresponse did not significantly differ between DQT and SET.

Interviewer training may help overcome subjective barriers of interviewers in asking sensitive questions. No such training is possible, however, for interviewees and, hence, their level of uneasiness seems to be a more important issue. According to the theoretical remarks in Section 1, SET can be seen as a device that aims at reducing the level of interviewee's 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 a scale from 1=not at all uneasy to 10=very uneasy, respondents were asked: "How uneasy did you personally feel in answering our questions about sexual experiences and behavior?" Table 9 presents the recoded frequency distributions for DQT and SET.

[11] As mentioned at the beginning of Section 2, the seminar was managed by the author and included 47 participating students who finally conducted the interviews.

Table 9 about here

Neither in DQT nor SET mode is the level of uneasiness very high. Nevertheless, the difference between DQT and SET percentages in Table 9 is highly significant in the predicted direction (chi-squared test). The same holds true if we do not recode the 10-digit scale and compare the medians and means which are 3.0 and 3.8 respectively for DQT and 2.0 and 2.9 for SET. Therefore, in line with expectations, SET reduces the level of interviewee's feelings of discomfort.[12]

[12] In this context, we also asked the question whether the respondent gave "more or less true" answers to the sex questions (10-digit scale from 1=not at all true to 10=completely true). Regrettably, this question did not show much variance. 79% of the DQT and 81% of the SET respondents did not hesitate to declare "completely true" (code 10). The DQT/SET differences are not significant here.

6. Conclusion

On the basis of the preceding results, we can conclude that – compared to direct questions – SET seems to have some merits: It reduces the underreporting bias, 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 contributes to decreasing item nonresponse; it lowers the level of uneasiness on the side of the interviewer and the interviewee; and – not previously mentioned – it brings in an element of variation into the interview making the interaction more interesting (at least for some respondents).

Of course, this study has limitations. The two most evident are: (1) It is confined to the area of sexual behavior and, thus, leaves out sensitive topics in other areas. (2) Its sample is restricted to university students and, thus, does not pertain to the general population. 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 is to be expected 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 shown in future work.

One general objection against SET could be that its merits simply reflect the advantages of self-administered mail or Internet surveys (greater belief in anonymity and confidentiality, no interviewer effects).[13] This is partly true, but to switch completely to a self-administered mail or Internet mode does not necessarily seem to be good advice for sensitive topics. A "sensitive topic survey" (like ours) via mail or Internet would probably have a low response rate, i.e. unit nonresponse would be high. Furthermore, via mail or Internet we would not have control over the response situation (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.

[13] With respect to sex questions, Tourangeau et al. (1997), for example, have demonstrated that self-administered questions (SAQs) produce higher levels of reporting.

References

Barnett, Julie, 1998: Sensitive Questions and Response Effects: An Evaluation. Journal of Managerial Psychology 13: 63-76.

Barton, Allen H., 1958: Asking the Embarrassing Question. Public Opinion Quarterly 22: 67-68.

Becker, Rolf and Ralph Günther, 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 28 (Nr. 54): 39-59.

Becker, Rolf and Ralph Günther, 2006: Selective Response to Questions on Delinquency. Quality & Quantity 40: 483-498.

Biemer, Paul P. and Lars E. Lyberg, 2003: Introduction to Survey Quality. Hoboken: Wiley-Interscience.

Bradburn, Norman M. and Seymour Sudman, 1979: Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass.

De Leeuw, Edith D., 2001: Reducing Missing Data in Surveys: An Overview of Methods. Quality & Quantity 35: 147-160.

Esser, Hartmut, 1986: Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als „rational choice“. Pp. 59-78 in: Reinhard Wittenberg (Ed.): Person – Situation – Institution – Kultur. Günter Büschges zum 65. Geburtstag. Berlin: Duncker & Humblot.

Hox, Joop J. and Edith D. de Leeuw, 2002: The Influence of Interviewers’ Attitude and Behavior on Household Survey Nonresponse: An International Comparison. Pp. 103-120 in: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (Eds.): Survey Nonresponse. New York: Wiley.

Jasso, Guillermina, 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: 224-241.

Krosnick, Jon A., et al., 2002: The Impact of “No Opinion” Response Options on Data Quality. Non-Attitude Reduction or an Invitation to Satisfice? Public Opinion Quarterly 66: 371-403.

Lee, Raymond M., 1993: Doing Research on Sensitive Topics. London: Sage.

Makkai, Toni and Ian Mcallister, 1992: Measuring Social Indicators in Opinion Surveys: A Method to Improve Accuracy on Sensitive Questions. Social Indicators Research 27: 169-186.

Perry, Paul, 1979: Certain Problems in Election Survey Methodology. Public Opinion Quarterly 43: 312-325.

Preisendörfer, Peter, 2008: Heikle Fragen in mündlichen Interviews: Ergebnisse einer Methodenstudie im studentischen Milieu. Mainz: Institute of Sociology, University of Mainz.

Schnell, Rainer, 1997: Nonresponse in Bevölkerungsumfragen. Opladen: Leske und Budrich.

Schnell, Rainer and Frauke Kreuter, 2005: Separating Interviewer and Sampling-Point Effects. Journal of Official Statistics 21: 389-410.

Smith, Tom W., 1992: Discrepancies Between Men and Women in Reporting Number of Sexual Partners: A Summary from Four Countries. Social Biology 39: 203-211.

Stocké, Volker, 2007a: The Interdependence of Determinants for the Strength and Direction of Social Desirability Bias in Racial Attitude Surveys. Journal of Official Statistics 23: 493-514.

Stocké, Volker, 2007b: Determinants and Consequences of Survey Respondents' Social Desirability Beliefs about Racial Attitudes. Methodology 3: 125-138.

Stocké, Volker and Christian Hunkler, 2007: Measures of Desirability Beliefs and Their Validity as Indicators for Socially Desirable Responding. Field Methods 19: 313-336.

Sudman, Seymour and Norman M. Bradburn, 1974: Response Effects in Surveys: A Review and Synthesis. Chicago: Aldine.

Sudman, Seymour and Norman M. Bradburn, 1982: Asking Questions: A Practical Guide to Questionnaire Design. San Francisco: Jossey-Bass.

Tourangeau, Roger, Kenneth A. Rasinski, Jared B. Jobe, Tom W. Smith and William F. Pratt, 1997: Sources of Error in a Survey on Sexual Behavior. Journal of Official Statistics 13: 341-365.

Tourangeau, Roger, Lance J. Rips and Kenneth A. Rasinski, 2000: The Psychology of Survey Response. Cambridge: Cambridge University Press.

Tourangeau, Roger and Ting Yan, 2007: Sensitive Questions in Surveys. Psychological Bulletin 133: 839-883.

Table 1: Sexual behavior comparing the direct question technique (DQT) and the sealed envelope technique (SET)

		DQT (n=211)	SET (n=367)	Significance of difference (* p < 0.05)
Sexual partners: no. in lifetime	Median:	4.0	4.0	
	Mean:	5.9	6.6	p=0.40
Coital frequency: no. in last 4 weeks	Median:	3.0	3.0	
	Mean:	4.7	5.0	p=0.62
One-night stand: ever		53.8%	49.3%	p=0.30
Sexual infidelity in current/ last partnership: ever		12.5%	10.5%	p=0.48
Homosexual contact: ever		14.8%	18.5%	p=0.26
Masturbation: often, very often		14.6%	29.4%	p=0.00*

Sexual partners: Question "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."

Coital frequency: Question "In the last four weeks, how many times have you had intercourse? If you do not remember this exactly, please try to give an estimation."

One-night stand: Question "Concerning your sexual experiences, have you ever had a one-night stand?" with answers yes/no.

Sexual infidelity: Question "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: Question "Have you ever had a homosexual relationship?" with answers yes/no.

Masturbation: Question "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.

Table 2: Factors affecting the number of sexual partners in a respondent's lifetime (OLS regression coefficients)

Independent variables	Model 1 ALL	Model 2a DQT	Model 2b SET	Significance of DQT-SET difference
SET (0/1)	0.02 (0.12)			
Female person (0/1)	0.36* (2.18)	0.43 (1.61)	0.34 (1.63)	p=0.78
Age in years	0.12* (5.40)	0.18* (5.12)	0.08* (3.03)	p=0.03*
Partnership (0/1)	0.63* (3.81)	0.61* (2.26)	0.62* (2.99)	p=0.96
Religious person (1-5)	-0.29* (4.17)	-0.45* (3.80)	-0.21* (2.41)	p=0.11
Party person (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	
Number of cases	555	196	359	

Notes: * $p < 0.05$; t-values in parentheses; reference groups: DQT, male, currently not living in a partnership; the dependent variable "number of sexual partners" has been transformed into logarithmic values.

Table 3: Factors affecting coital frequency in the last four weeks (OLS regression coefficients)

Independent variables	Model 1 ALL	Model 2a DQT	Model 2b SET	Significance of DQT-SET difference
SET (0/1)	-0.04 (0.16)			
Female person (0/1)	-0.25 (0.95)	-0.05 (0.12)	-0.39 (1.14)	p=0.53
Age in years	-0.01 (0.29)	-0.02 (0.42)	-0.01 (0.07)	p=0.79
Partnership (0/1)	5.67* (21.59)	6.22* (15.59)	5.34* (15.49)	p=0.11
Religious person (1-5)	-0.21 (1.90)	-0.28 (1.62)	-0.17 (1.16)	p=0.63
Party person (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	
Number of cases	557	200	357	

Notes: * $p < 0.05$; t-values in parentheses; reference groups: DQT, male, currently not living in a partnership; the dependent variable "coital frequency" has been transformed into logarithmic values.

Table 4: Factors affecting the probability of one-night stands (binary logistic regression coefficients)

Independent variables	Model 1 ALL	Model 2a DQT	Model 2b SET	Significance of DQT-SET difference
SET (0/1)	-0.16 (0.89)			
Female person (0/1)	0.12 (0.66)	0.24 (0.79)	0.05 (0.24)	p=0.62
Age in years	0.07* (2.92)	0.15* (3.02)	0.04 (1.25)	p=0.05
Partnership (0/1)	-0.14 (0.81)	-0.58 (1.88)	0.09 (0.40)	p=0.08
Religious person (1-5)	-0.26* (3.36)	-0.37* (2.76)	-0.21* (2.21)	p=0.33
Party person (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)	
Pseudo R^2	0.04	0.09	0.03	
Number of cases	568	207	361	

Notes: * $p < 0.05$; t-values in parentheses; reference groups: DQT, male, currently not living in a partnership; the dependent variable is whether the respondent ever had a one-night stand.

Table 5: Factors affecting sexual infidelity in partnerships (binary logistic regression coefficients)

Independent variables	Model 1 ALL	Model 2a DQT	Model 2b SET	Significance of DQT-SET difference
SET (0/1)	-0.16 (0.56)			
Female person (0/1)	-0.31 (1.13)	-0.27 (0.62)	-0.36 (0.99)	p=0.87
Age in years	0.01 (0.03)	-0.05 (0.77)	0.03 (0.65)	p=0.31
Partnership (0/1)	-0.27 (0.97)	-0.35 (0.80)	-0.24 (0.66)	p=0.83
Religious person (1-5)	-0.06 (0.47)	-0.10 (0.47)	-0.04 (0.27)	p=0.84
Party person (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)	
Pseudo R^2	0.01	0.02	0.01	
Number of cases	540	199	341	

Notes: * $p < 0.05$; t-values in parentheses; reference groups: DQT, male, currently not living in a partnership; the dependent variable is whether the respondent was sexually unfaithful in his/her current or last partnership.

Table 6: Factors affecting the probability of homosexual contact (binary logistic regression coefficients)

Independent variables	Model 1 ALL	Model 2a DQT	Model 2b SET	Significance of DQT-SET difference
SET (0/1)	0.27 (1.08)			
Female person (0/1)	1.08* (4.05)	1.39* (2.83)	0.97* (3.03)	p=0.47
Age in years	-0.02 (0.65)	-0.07 (0.93)	-0.01 (0.39)	p=0.52
Partnership (0/1)	0.62* (2.49)	-0.06 (0.15)	0.97* (2.97)	p=0.05*
Religious person (1-5)	-0.15 (1.49)	-0.49* (2.40)	-0.04 (0.30)	p=0.06
Party person (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)	
Pseudo R^2	0.06	0.09	0.07	
Number of cases	572	209	363	

Notes: * $p < 0.05$; t-values in parentheses; reference groups: DQT, male, currently not living in a partnership; the dependent variable is whether the respondent has ever had homosexual contact.

Table 7: Factors affecting masturbation frequency (OLS regression coefficients)

Independent variables	Model 1 ALL	Model 2a DQT	Model 2b SET	Significance of DQT-SET difference
SET (0/1)	0.34* (4.06)			
Female person (0/1)	-0.75* (9.22)	-0.45* (3.59)	-0.91* (8.74)	p=0.01*
Age in years	0.01 (0.63)	0.02 (1.44)	-0.01 (0.35)	p=0.19
Partnership (0/1)	-0.16 (1.95)	-0.34* (2.71)	-0.06 (0.61)	p=0.10
Religious person (1-5)	-0.11* (3.16)	-0.20* (3.66)	-0.07 (1.56)	p=0.07
Party person (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	
Number of cases	553	198	355	

Notes: * $p < 0.05$; t-values in parentheses; reference groups: DQT, male, currently not living in a partnership; the dependent variable "masturbation frequency" is measured on a 5-digit scale from 1=never to 5=very often.

Table 8: Missing values of the sexual behavior questions comparing the direct question technique (DQT) and the sealed envelope technique (SET)

	DQT (n=211)	SET (n=367)	Significance of difference (* p < 0.05)
Sexual partners: no. in lifetime	6.6% (n=14)	1.6% (n=6)	p=0.00*
Coital frequency: no. in last 4 weeks	4.7% (n=10)	1.9% (n=7)	p=0.05
One-night stand: ever	1.4% (n=3)	1.1% (n=4)	p=0.72
Sexual infidelity in current/ last partnership: ever	0.5% (n=1)	0.0% (n=0)	p=0.19
Homosexual contact: ever	0.5% (n=1)	0.8% (n=3)	p=0.63
Masturbation: often, very often	5.7% (n=12)	2.7% (n=10)	p=0.07
Average no. of missing values for the 6 items	0.19	0.08	p=0.02*

Table 9: Level of respondents' uneasiness of the sexual behavior questions comparing the direct question technique (DQT) and the sealed envelope technique (SET)

Level of respondents' uneasiness	DQT (n=211)	SET (n=367)	Significance of difference (* p < 0.05)
Low (codes 1-3)	52.6%	67.6%	
Moderate (codes 4-7)	37.4%	28.0%	p=0.00*
High (codes 8-10)	10.0%	4.4%	

Question: "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.