

ASKING SENSITIVE QUESTIONS USING THE CROSSWISE MODEL AN EXPERIMENTAL SURVEY MEASURING PLAGIARISM

BEN JANN*

JULIA JERKE

IVAR KRUMPAL*

Abstract Yu, Tian, and Tang (2008) proposed two new techniques for asking questions on sensitive topics in population surveys: the triangular model (TM) and the crosswise model (CM). The two models can be used as alternatives to the well-known randomized response technique (RRT) and are meant to overcome some of the drawbacks of the RRT. Although Yu, Tian, and Tang provide a promising theoretical analysis of the proposed models, they did not test them. We therefore provide results from an experimental survey in which the crosswise model was implemented and compared to direct questioning. To our knowledge, this is the first empirical evaluation of the crosswise model. We focused on the crosswise model because it seems better suited than the triangular model to overcome the self-protective “no” bias observed for the RRT. This paper-and-pencil survey on plagiarism was administered to Swiss and German students in university classrooms. Results suggest that the CM is a promising data-collection instrument eliciting more socially undesirable answers than direct questioning.

BEN JANN is an Associate Professor at the Institute of Sociology, University of Bern, Bern, Switzerland. JULIA JERKE is a graduate student at the Institute of Sociology, University of Leipzig, Leipzig, Germany. IVAR KRUMPAL is a senior researcher at the Institute of Sociology, University of Leipzig, Leipzig, Germany. The authors thank Debra Hevenstone and the anonymous reviewers for their comments on an earlier draft of this article. Furthermore, the authors thank Norman Braun, Jochen Groß, and Matthias Naef for their assistance with the data collection. This work was supported by the German Research Foundation [DI 292/5 to Andreas Diekmann, VO 684/11 to Thomas Voss and Karl-Dieter Opp]. *Address correspondence to Ben Jann, University of Bern, Institute of Sociology, Lerchenweg 36, 3000 Bern 9, Switzerland; e-mail: jann@soz.unibe.ch; or Ivar Krumpal, University of Leipzig, Institute of Sociology, Beethovenstrasse 15, 04107 Leipzig, Germany; e-mail: krumpal@sozio.uni-leipzig.de.

Introduction

It is challenging to accurately measure sensitive issues such as personal, illegal, or socially undesirable behavior because respondents tend to conceal offenses against the law and deny violations of social norms. Such systematic response errors lead to social-desirability bias in prevalence estimates of the sensitive behaviors of interest, underestimating socially undesirable activities such as abortion, illicit drug use, social fraud, and student cheating, and overestimating socially desirable activities such as voting, seat belt use, and exercising (for an overview, see Tourangeau and Yan 2007). Another severe problem is item non-response if sensitive questions are asked directly. Some respondents may refuse to reveal personal information about their income or sexual behavior if they perceive such questions as too intrusive or embarrassing. If item non-response is systematically related to the behaviors of interest, estimates will be distorted.

THE RANDOMIZED RESPONSE TECHNIQUE

To overcome these problems, methods such as the randomized response technique (RRT) have been developed. In the RRT, respondents use a randomization device with a known probability distribution (e.g., coins, dice, cards) to generate a probabilistic relationship between their answers and the true values of the sensitive characteristic. In Warner's original RRT design (Warner 1965), respondents are confronted with two statements, *A* and *B*, one of which is the negation of the other. For example, the two statements might be "I have cheated on a written test at least once" (statement *A*) and "I have never cheated on a written test" (statement *B*). Respondents are then asked to employ a randomization device to pick one of the two statements and indicate whether the randomly chosen statement is true or not. For example, a respondent may be instructed to throw a die (unobserved by the interviewer) and answer statement *A* if the outcome is 1 or 2 and statement *B* if the outcome is 3, 4, 5, or 6. The meaning of an observed "yes" or "no" answer depends on the die's result, which is known only to the respondent. Therefore, a given answer does not reveal anything definite about the true state of the sensitive characteristic. Being aware of this anonymity, respondents are expected to give more honest answers than if asked using non-anonymous direct questioning. However, if the probabilities of statements *A* and *B* are known, an estimator for the prevalence of the sensitive characteristic in the population can be derived, even though no direct link between the observed answers and the variable of interest exists on the individual level.

The principle of adding random noise to the data to establish a probabilistic link between the observed answers and the sensitive question is also at the heart of alternative RRT schemes (for an overview, see Fox and Tracy 1986). A popular variant is Boruch's (1971) forced-choice method that requires the

respondent to either answer the sensitive question (e.g., statement A in the example above) or give an automatic “yes” or “no” answer depending on the outcome of the randomization device. For example, respondents could be instructed to flip two coins and answer “yes” if the result is heads for both coins and “no” in the case of tails for both coins. Only if the results for the two coins differ is the respondent instructed to answer the sensitive question. Again, because the coins’ results are private information, a “yes” answer does not reveal the true state of the sensitive characteristic. At the same time, however, estimation of the population prevalence is possible if the probabilities of the different outcomes of the randomization device are known. Regression analysis of covariate effects is also possible based on RRT data (Maddala 1983; Scheers and Dayton 1988; van der Heijden et al. 2000; Jann 2005).

Although the RRT is straightforward in principle, there are a number of practical problems that complicate its application. For the RRT estimates to be valid, it has to be assumed that the respondents comprehend the somewhat complex RRT procedure, appreciate the induced anonymity, and are ready to answer the sensitive question truthfully if instructed to do so by the protocol. In other words, they need to understand and trust in the RRT scheme.¹ Studies show that these assumptions are often violated (Cruyff et al. 2007; Landsheer, van der Heijden, and van Gils 1999; Lensvelt-Mulders and Boeije 2007; Coutts and Jann 2011). A substantial proportion of respondents, misunderstanding or mistrusting the RRT, may not comply with the RRT instructions or may refuse to reply. Since both types of respondents—those who *did* commit the sensitive behavior as well as those who *did not*—can introduce self-protective response bias in the RRT, the overall potential for social-desirability bias is even higher than for direct questioning. For example, in the forced-choice design, both types of respondents are tempted to deviate from the RRT protocol and provide a self-protective “no” answer even if the outcome of the randomization device requires them to say “yes.” A further problem of the RRT is that a randomization device has to be employed (coins, dice, colored marbles, etc.; for a list, see Lensvelt-Mulders et al. 2005, p. 335). This imposes an additional burden on the respondents and, especially in self-administered modes, respondents may choose to skip randomization or even break off the interview. Finally, extensive interviewer training (when the survey is being administered by an interviewer) and sophisticated instructions (when the survey is self-administered) are required to gain respondents’ trust

1. For example, a common misunderstanding in Warner’s model is that the truth is revealed with a certain probability (Warner 1986, p. 441): “For those not familiar with probability, requiring the ‘actual answer’ or ‘actual value’ to be reported with probability P may be misinterpreted as requiring that the secret which the interviewee wishes to hide will be revealed with probability P .” Another difficulty is the occurrence of double negatives in Warner’s RRT scheme (e.g., exam cheaters who have to answer “no” to “I have never cheated on a written test”). Double negations increase the cognitive burden for respondents and lead to erroneous answers.

and avoid misunderstanding of the RRT procedure. This requires extra pre-testing and increases the costs of data collection.

THE TRIANGULAR MODEL AND THE CROSSWISE MODEL

To address the problem of self-protective response bias, statistical methods have been proposed to estimate the proportion of respondents who do not comply with the RRT instructions and to correct the RRT estimates accordingly (Ostapczuk, Musch, and Moshagen 2009). Alternatively, Yu, Tian, and Tang (2008) recently proposed two new data-collection methods, the triangular model (TM) and the crosswise model (CM), which may overcome some of the drawbacks of the RRT. According to Yu, Tian, and Tang (2008, pp. 261–62), the two models are attractive for both face-to-face and self-administered interviews because no randomization device is required and the practical implementation is less complex than for the RRT. A further advantage is that none of the respondents has to answer the sensitive questions directly. In contrast, RRT schemes require at least some of the respondents to provide a direct answer to the sensitive question.

Both techniques, the TM and the CM, are based on a simple idea (Yu, Tian, and Tang 2008, pp. 254–55): Respondents are exposed to two questions, a sensitive one and a non-sensitive one, and are asked to provide a joint answer to both questions. Let X be the sensitive variable of interest. X is equal to 1 if the sensitive behavior applies, e.g., if the respondent cheated on an exam, and 0 otherwise. Furthermore, let Y be a non-sensitive variable that is unrelated to X and for which the probability distribution is known. For example, let Y be equal to 1 for people born in October, November, or December, and 0 otherwise. Respondents are then asked to provide a joint answer for X and Y by choosing one of two response options, either A or B , according to a specific rule.

In the triangular model (TM), respondents choose option A if the answer is “no” to both questions and option B if at least one answer is “yes.” Hence, option A refers to the subgroup of “innocent” respondents who did not cheat on an exam and were born between January and September. Option B , on the other hand, applies to respondents who either cheated or were born between October and December (or both). Formally, A refers to subgroup $\{X = 0 \cap Y = 0\}$, whereas B refers to subgroup $\{X = 1 \cap Y = 1\} \cup \{X = 1 \cap Y = 0\} \cup \{X = 0 \cap Y = 1\}$. Because the latter contains a mixture of “innocent” and “guilty” respondents, choosing option B cannot be interpreted as an admission of guilt. Yu, Tian, and Tang (2008) conjecture that respondents understand the anonymity the design generates and are therefore more willing to provide honest answers. However, similar to the forced-choice RRT, an obvious self-protective answer strategy exists for the TM: Choosing option A unambiguously indicates that the individual did not engage in the

sensitive behavior. Option *A* is therefore the dominant strategy for respondents who do not want to have any association with the sensitive behavior.

The crosswise model (CM) might overcome this drawback. The response rule for the CM is to choose option *A* if the answer is the same for both questions (both “yes” or both “no”) and option *B* if one answer is “yes” and the other is “no” (for example, see our implementation of the CM in appendix A). That is, *A* now refers to subgroup $\{X = 1 \cap Y = 1\} \cup \{X = 0 \cap Y = 0\}$ and *B* refers to $\{X = 1 \cap Y = 0\} \cup \{X = 0 \cap Y = 1\}$. Since both groups contain a mixture of “guilty” and “innocent” respondents, there is no longer an obvious self-protective answer strategy. We therefore expect that the CM leads to more valid measurement than does the TM. Note that in the CM, the (known) probability distribution for the non-sensitive variable must be uneven; that is, $\Pr(Y = 1)$ must not be equal to 0.5.

OUR STUDY

Although Yu, Tian, and Tang (2008) provide a promising analysis of the theoretical properties of the proposed models (i.e., derivation of unbiased maximum likelihood estimates and asymptotic properties, optimization of design parameters and sample-size formulas, efficiency analyses based on the variance criterion), it remains unclear how the techniques will perform in empirical practice, as Yu, Tian, and Tang (2008) did not test their proposal. We provide results from an experimental survey in which one of the techniques, the CM, was implemented and compared to direct questioning. To our knowledge, this is the first empirical application and evaluation of the CM. We focus on the CM because it seems better suited than the TM to overcome self-protective answer bias. We apply the technique to measure the prevalence of plagiarism in student papers. In addition to prevalence estimates, we also report results from adapted regression models to analyze the relationship between plagiarism and covariates based on CM data.

Plagiarism in Student Papers

According to the Office of Science and Technology Policy, plagiarism is the “appropriation of another person’s ideas, processes, results, or words without giving appropriate credit” (2000, p. 76,262). The term “plagiarism” also includes copying or paraphrasing a text passage from someone else’s work without citing the original (Park 2003, p. 472). In the age of the Internet and Wikipedia, universities increasingly worry about plagiarism in student papers and homework assignments and have consequently issued guidelines for lecturers and students that define plagiarism and explain the consequences of plagiarism. For example, the disciplinary code of the ETH Zurich (ETH Zurich Executive Board 2004) specifies that the code is violated if a student

“hands in a written assignment that he/she has not written himself/herself, or in which he/she passes off as one’s own the results and insights of another (plagiarism)” (Art. 2b). Depending on the type and severity of the offense, different sanctions can be imposed, such as issuing a reprimand; declaring a student’s performance assessments as failed; threatening to or actually suspending the person from specific courses, facilities, or the university for a maximum of three years; or divesting the person of an academic title if acquired illicitly (Art. 3).

Plagiarism is socially undesirable behavior that can be severely sanctioned. Being aware of the risks and negative repercussions associated with cheating, plagiarists probably purport to be honest students who submit original work. For universities, however, it might be important to know how frequently different forms of plagiarism occur.

Valid prevalence estimates of plagiarism in student papers are difficult to obtain. The official number of students found guilty is a poor indicator of the true level of plagiarism because many cases may remain undetected. An alternative approach to estimate the prevalence of plagiarism is to systematically check a sample of student papers using Internet search engines (Krohn, Schlombs, and Taubert 2003) or specialized software such as “Turnitin” or “Plagiarism-Finder” (Sattler 2007). Inspecting student papers has the advantage of being a nonreactive method to collect data and estimate the prevalence of plagiarism without relying on self-reports. However, software-based methods may be inaccurate. First, not all incidences of plagiarism are found (i.e., there are “false negatives”). Second, diagnosed non-originality usually also includes cases where material has been legitimately cited (“false positives”). Culwin (2006, p. 169) concludes that, overall, “the levels of non-originality reported from such studies can be regarded as a low-water mark for the true extent of non-originality.”

An alternative approach is to use surveys (Marsden, Carroll, and Neill 2005). Because plagiarism is a highly sensitive issue, however, it is expected that many students will misreport their behavior and provide answers that are biased toward conformity with the norms and regulations of their university. As argued above, the proposed CM appears to be a promising data-collection method suitable to overcome this deficiency. Compared to direct questioning, the CM is assumed to generate a higher sense of protection and is thus expected to yield more accurate estimates of plagiarism.

Study Design

To evaluate the viability of the crosswise model (CM), we employed a randomized experimental design using two different questionnaires: a direct-questioning version and a CM version. In accordance with the “more-is-better” assumption (see Lensvelt-Mulders et al. 2005; Krumpal

2008), we expect higher prevalence estimates for the CM than for direct questioning if the CM is more successful in eliciting honest answers. Except for the differing experimental part, the two questionnaires were identical. After a short introductory text, a few questions about personal characteristics and field of study, and some general questions about writing student papers (e.g., number of papers written, information sources), the sensitive questions related to plagiarism were asked. We asked two sensitive questions, one about partial plagiarism and one about severe plagiarism.² The wording of the questions was (translated from German):

- “When writing an assignment (e.g., seminar paper, term paper, thesis), have you ever intentionally adopted a passage from someone else’s work without citing the original?” (partial plagiarism)
- “Did you ever have someone else write a large part of an assignment for you or hand in someone else’s work (e.g., from www.hausarbeiten.de) as your own?” (severe plagiarism)

In the direct-questioning condition, students had to answer the two sensitive questions directly. In the CM condition, the sensitive questions were paired with two non-sensitive items and students were asked to provide a joint answer to each pair. The two non-sensitive items were as follows:

- “Is your mother’s birthday in January, February, or March?” (paired with the partial plagiarism question)
- “Is your father’s birthday in October, November, or December?” (paired with the severe plagiarism question)

For the CM estimator to be unbiased and identified, the non-sensitive items have to be uncorrelated with the sensitive questions and the probability of a “yes” answer has to be known and unequal to 0.5. The non-sensitive birthday questions are likely uncorrelated with plagiarism, and the probability of a positive answer is about 0.25, assuming a uniform distribution of birthdays.

2. To minimize the possibility of context effects, we positioned the experimental section containing the sensitive questions at the end of the questionnaire. Hence, we think it is safe to assume that the experimental manipulation did not have a substantial effect on how the other questions were answered. However, the cognitive burden imposed on the respondents is higher for the CM than for direct questioning, which might result in a higher amount of random response error in the CM condition. This should not affect the prevalence estimates, but it might blur the associations with other variables.

The CM is more difficult to communicate to the respondents than direct questioning, especially in a self-administered survey. We therefore developed suitable instructions based on pretests including cognitive interviews.³ The final implementation of the CM used in our study is illustrated in appendix A (translated from German).

Our survey was conducted between June and July 2009 among 474 German and Swiss university students. The questionnaires were distributed in lectures at the University of Leipzig ($n = 273$), the ETH Zurich ($n = 111$), and the LMU Munich ($n = 90$). In each class, students were randomly assigned to one of two experimental conditions, using a ratio of 3 for the CM ($n = 358$) to 1 for direct questioning ($n = 116$).⁴ The CM was oversampled because it has lower statistical power than direct questioning. A χ^2 -test of independence between experimental condition and university confirms the null hypothesis ($\chi^2 = 0.094$, $p = 0.954$). Net of the students who did not write any papers yet and net of item non-response, which was as low as 1 percent for both experimental conditions (1 missing value for each item in the direct-questioning condition and 3 missing values for each item in the CM), there are 310 valid observations for the CM and 96 valid observations for direct questioning.⁵

3. After developing a first version of the questions and the CM instructions, we presented the questionnaire to research colleagues at our universities (all experienced social scientists and survey researchers). Based on these expert reviews, the CM instructions were simplified and ambiguous formulations were clarified. Next, we performed cognitive interviews with a convenience sample of 10 undergraduate students from the University of Leipzig (seven students for the CM version of the questionnaire, three students for the direct-questioning version). Retrospective-think-aloud was used to evaluate the questionnaires; that is, after each block of questions (socio-demographic variables, general questions about writing an assignment, the sensitive questions), the respondents were asked to verbalize their thoughts. For the assessment of the CM instructions, we focused on (1) whether respondents understood the instructions and were able to apply them to the specific question-and-answer task; and (2) whether respondents trusted the privacy protection. None of the respondents showed notable difficulties with respect to the first criterion. However, four respondents expressed mistrust regarding the privacy protection. Follow-up questions in which we asked the respondents to specify their concerns revealed that a sentence containing a statement about the use of the “known birthday distribution” for computing the prevalence rate was responsible for the mistrust. The statement suggested that something could be “known,” thus reducing the perceived anonymity. To avoid this, we removed the statement from the instructions.

4. The classroom setting facilitated cooperation. We did not encounter any open objections to the survey, and no empty questionnaires were returned. (To motivate participation, the scientific character of the study was emphasized and, more importantly, the students were given time to complete the questionnaire. No monetary incentives were provided.) Nonetheless, the sample cannot be considered representative of any well-defined population because the classes were not randomly selected and probably not all enrolled students were present at the day of the data collection.

5. Sixty-five students stated that they had not written any papers. One of them nonetheless completed the follow-up questions and also gave an answer to the sensitive items. Assuming that this student erroneously checked a wrong answer in the filter question, we decided to include this observation in our analyses.

Table 1. Frequency Distributions for Age, Sex, and Nationality by Experimental Condition

	Experimental Condition		Chi-Squared
	Direct Questioning	Crosswise Model	
Age			
18 to 22	49.5%	49.5%	1.31
23 to 27	44.2%	40.5%	(p = 0.521)
28 or older	6.3%	10.0%	
Sex			
Female	56.7%	50.3%	1.20
Male	43.3%	49.7%	(p = 0.273)
Nationality			
German or Swiss	95.9%	92.9%	1.09
Other	4.1%	7.1%	(p = 0.296)

Table 1 displays the distribution of some socio-demographic characteristics of these respondents. The average age of the participants was 23.2 years, and about half of them were female (51.8 percent overall). The nationality of the vast majority of respondents was German or Swiss (93.6 percent). χ^2 -tests of independence provide no evidence for systematic differences in the distributions of the socio-demographic characteristics between the two experimental conditions.

Results

Prevalence estimation for direct questioning is straightforward. For the CM, an estimator can be derived as follows. Let X denote the sensitive item with unknown population prevalence $\pi = \Pr(X = 1)$. Furthermore, let Y be the non-sensitive item with known population prevalence $p = \Pr(Y = 1)$. In our study, $p = 0.25$. Given that X and Y are uncorrelated, that is, $\text{Cov}(X, Y) = 0$, the probability of answer option A in the CM ($X = 1$ and $Y = 1$ or $X = 0$ and $Y = 0$) is $\phi = p\pi + (1 - p)(1 - \pi)$. Thus, a natural estimator for π is

$$\hat{\pi}_{\text{CM}} = \frac{(\hat{\phi} + p - 1)}{(2p - 1)}, \quad p \neq 0.5,$$

where $\hat{\phi}$ is the observed proportion of respondents choosing response option A (i.e., “yes” to both or “no” to both the sensitive and the non-sensitive question).⁶ Furthermore, the sampling variance of $\hat{\pi}_{\text{CM}}$ can be estimated as

6. Note that the CM estimator is formally identical to the estimator for Warner’s original RRT design (Warner 1965; Yu, Tian, and Tang 2008).

$$\text{Var}(\hat{\pi}_{\text{CM}}) = \frac{\hat{\phi}(1 - \hat{\phi})}{n(2p - 1)^2} = \frac{\hat{\pi}_{\text{CM}}(1 - \hat{\pi}_{\text{CM}})}{n} + \frac{p(1 - p)}{n(2p - 1)^2}.$$

The plagiarism prevalence estimates are presented in table 2. For the partial plagiarism item, a statistically significant difference of about 15 percentage points can be observed between the two measurement methods ($p = 0.014$). According to direct questioning, the prevalence of partial plagiarism is 7.3 percent, compared to 22.3 percent for the CM. For partial plagiarism, we can therefore confirm the expectation that the CM yields higher prevalence estimates than direct questioning. In contrast, however, the estimates for severe plagiarism do not seem to differ between the two measurement methods. The estimates are 1 percent for direct questioning and 1.6 percent for the CM. Although the difference is substantial in relative terms, the two estimates cannot be distinguished statistically due to a lack of precision in estimating probabilities that are close to 0 (or 1). This is especially true for the CM for which the standard error decreases only marginally, as the target probability moves toward the boundaries of the probability space.

In addition to prevalence estimation, the analysis of effects of covariates on the probability of plagiarism may be of interest. Standard binary-response models cannot be used to analyze CM data because the observed response is a mixture of the variable of interest, X , and the auxiliary variable, Y . However, since the probability distribution of Y is known (and the two variables are independent), it is possible to formulate adapted regression estimators for CM data. Two such estimators are a modified logistic regression model and a modified linear probability model (LPM).

Let $\pi = \Pr(X = 1|Z)$ be the (unknown) probability of a “yes” answer to the sensitive question conditional on a vector of (observed) covariates Z (including a constant). In logistic regression, π is linked to covariates using the logistic function; that is,

$$\pi = \frac{e^{Z'\beta}}{1 + e^{Z'\beta}},$$

where β denotes an unknown parameter vector. Standard maximum likelihood procedures can be used to estimate β if X is observed. In particular, in a simple

Table 2. Prevalence Estimates, in Percent, of Partial and Severe Plagiarism by Experimental Condition (standard errors in parentheses)

	Direct Questioning (DQ)	Crosswise Model (CM)	Difference (CM – DQ)
Partial Plagiarism	7.3 (2.7)	22.3 (5.5)	15.0 (6.1)
Severe Plagiarism	1.0 (1.0)	1.6 (5.0)	0.6 (5.1)
Observations	96	310	

random sample, the probabilities of observing $X = 1$ and $X = 0$ are π and $1 - \pi$, respectively. Hence, given independent observations $i = 1, \dots, n$, the log likelihood for β is

$$\ln L(\beta|X, Z) = \sum_{i=1}^n [X_i \ln(\pi_i) + (1 - X_i) \ln(1 - \pi_i)] \text{ with } \pi_i = \frac{e^{Z_i' \beta}}{1 + e^{Z_i' \beta}}.$$

In the CM, however, X is unknown. Instead, a response variable R , defined as

$$R = \begin{cases} 1 & \text{if response is A } (X = 1 \text{ and } Y = 1 \text{ or } X = 0 \text{ and } Y = 0) \\ 0 & \text{if response is B } (X = 1 \text{ and } Y = 0 \text{ or } X = 0 \text{ and } Y = 1) \end{cases},$$

is observed. Let $p = \Pr(Y = 1|Z)$ be the (known) probability of a “yes” answer to the auxiliary question ($p \neq 0.5$). The probabilities of $R = 1$ and $R = 0$ are then equal to $\pi p + (1 - \pi)(1 - p)$ and $\pi(1 - p) + (1 - \pi)p$, respectively. Thus, the log likelihood for β can be written as

$$\ln L(\beta|R, Z) = \sum_{i=1}^n \ln l_i,$$

with

$$\begin{aligned} \ln l_i &= R_i \cdot \ln[\pi_i p_i + (1 - \pi_i)(1 - p_i)] + (1 - R_i) \cdot \ln[\pi_i(1 - p_i) + (1 - \pi_i)p_i] \\ &= R_i \cdot \ln[p_i e^{Z_i' \beta} + (1 - p_i)] + (1 - R_i) \cdot \ln[(1 - p_i) e^{Z_i' \beta} + p_i] - \ln[1 + e^{Z_i' \beta}]. \end{aligned}$$

Maximizing the log likelihood yields the CM logit estimate of β .⁷ The CM logit coefficients can be interpreted just like the coefficients from a regular logit model (see, e.g., Long 1997, ch. 3). A key difference, however, is that the CM logit is less efficient than the regular logit model (unless $p = 0$). That is, to obtain regression analyses with comparable statistical power, a larger sample is required for the CM than for direct questioning. Note that for $p = 0$, the CM logit reduces to the regular logit model. Therefore, data from the CM and from direct questioning can be jointly analyzed within the same model through observation-specific values for p .

An alternative approach to analyzing CM data is to assume a linear probability model (LPM) with $\Pr(X = 1|Z) = Z' \beta$ and apply the least-squares method to a transformed response variable $\tilde{R} = (R + p - 1)/(2p - 1)$.⁸ The transformation follows from

$$\begin{aligned} E(R = 1|Z) &= \Pr(R = 1|Z) = (Z' \beta)p + (1 - Z' \beta)(1 - p) \\ &= (Z' \beta)(2p - 1) + 1 - p, \end{aligned}$$

7. The model can be easily adapted to allow for unequal selection probabilities, stratification, or multi-stage sampling employing standard methodology for complex survey data.

8. We use heteroscedasticity-robust formulas for the standard errors because the assumption of constant residual variance is violated by design.

and therefore

$$E[(R + p - 1)/(2p - 1)|Z] = \frac{E(R = 1|Z) + p - 1}{2p - 1} = Z'\beta,$$

where $E(.)$ denotes the expected value. The linearity assumption implied by the LPM may not be realistic. Nonetheless, the LPM is useful in practice because the resulting coefficient estimates can be interpreted as marginal effects on the probability of $X = 1$ without the need for further transformation. Moreover, in

Table 3. Results from Least-Squares and Maximum-Likelihood Regression Models of Partial Plagiarism (robust standard errors in parentheses)

	LPM (least-squares)		Logit (maximum-likelihood)	
	CM only	CM and DQ	CM only	CM and DQ
Crosswise model		0.17** (0.06)		1.64** (0.58)
University (Reference = LMU Munich)				
ETH Zurich	0.33+ (0.19)	0.26+ (0.15)	1.52 (1.19)	0.99 (0.98)
University of Leipzig	0.20 (0.14)	0.13 (0.11)	1.00 (0.92)	0.23 (0.76)
Female	0.03 (0.12)	0.04 (0.09)	0.23 (0.81)	0.52 (0.60)
Number of papers written (Reference = One or two)				
Three or four	-0.11 (0.14)	-0.10 (0.11)	-0.68 (0.86)	-0.80 (0.72)
Five or more	0.09 (0.14)	0.05 (0.11)	0.22 (0.83)	-0.34 (0.70)
Used information sources				
Internet	0.17 (0.17)	0.15 (0.14)	1.01 (1.59)	1.39 (1.85)
Other students' papers	0.19 (0.14)	0.16 (0.10)	0.96 (0.79)	1.07+ (0.56)
Constant	-0.17 (0.19)	-0.25 (0.16)	-3.30* (1.55)	-4.76** (1.74)
Observations	306	402	306	402

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

most cases, the average marginal effects from a logit model are very close to the coefficients from the LPM (see, e.g., Mood 2010).

Table 3 displays the results from logit and LPM models for our data. Partial plagiarism is regressed on variables such as university (University Leipzig, LMU Munich, ETH Zurich), gender, number of papers written, and whether the Internet and other students' papers were used as a source of information for writing papers. The models in columns one and three are based on CM data only; in columns two and four, data from the CM and from direct questioning are analyzed jointly. Models one and two are least-squares linear probability models; models three and four are maximum-likelihood logit models. Given the small number of observations and the low efficiency of the CM estimator, only a few statistically significant results can be observed. Plagiarism seems to be slightly more prevalent in Zurich than in Leipzig or Munich (the difference between Zurich and Munich is significant at the 10-percent level for the LPMs). No difference exists between males and females, and the number of papers written does not seem to be related to plagiarism (all p -values larger than 0.1). Furthermore, although all coefficients for the use of the Internet and other students' papers as information sources are positive, only one of them reaches the 10-percent significance level. The only clear effect we find is the effect of the questioning technique (see the coefficients for "Crosswise model" in the first row of table 3). According to the LPM, the rate of (admitted) partial plagiarism increases by 17 percentage points if the CM is used instead of direct questioning ($p = 0.007$). The logit model reveals a similar positive effect of the CM ($p = 0.005$). Models for severe plagiarism are not reported because the estimated prevalence is too low to be analyzed successfully by multivariate methods given such a small sample.

Conclusions

Our study investigated the viability of the CM with respect to the measurement of socially undesirable behavior in sensitive surveys. Adopting an experimental design, the performance of the CM was empirically evaluated and compared to conventional direct questioning. Our results show that underreporting of (partial) plagiarism is significantly reduced when the data are collected via the CM, and we thus conclude that the CM is successful in decreasing the social-desirability bias in prevalence estimates of plagiarism. Survey designers who aim to estimate the prevalence of sensitive behaviors could therefore benefit from using the CM in their research studies.

The strength of the CM, we believe, comes from the fact that, other than in the various implementations of the RRT, there is no condition in which respondents have to provide an immediate answer to the sensitive question, and there is no obvious self-protective answering strategy. However, further research to replicate our findings and extend the evaluation of the CM in other contexts and

settings would be desirable. For example, it would be worthwhile to implement and evaluate the CM in computer-assisted telephone interviews (CATI), in which information processing may be more difficult.

Furthermore, the CM could be compared to the triangular model (TM), which was also proposed by Yu, Tian, and Tang (2008) as an alternative question technique for asking sensitive questions. Likewise, it would be interesting to compare the CM to a classic RRT approach. Because there is no obvious self-protective answering strategy in the CM, one could hypothesize that the CM elicits more socially undesirable answers than the TM or RRT. We would expect to observe a larger difference between the CM and direct questioning than between the TM or RRT and direct questioning. Another method that might be less prone to self-protective response bias than the RRT is the item count technique (ICT; also known as the unmatched count technique or the list experiment). For example, see Holbrook and Krosnick (2010) or Coutts and Jann (2011) for promising results. The CM and ICT seem to have similar advantages over the RRT in terms of ease of use and self-protective response bias, and it would be valuable to contrast the two in an empirical study. We expect the CM to be superior to the ICT in terms of both answering time and statistical efficiency.

Finally, we have to note several limitations of our research and the CM. First, the present study is based on a relatively small sample of university students. The CM proved useful in reducing social-desirability bias in this small sample. However, we have to be careful not to prematurely generalize this result to other population groups. More empirical evidence on the viability of the CM is needed before this technique can be advocated as a generally effective instrument for reducing social-desirability bias. In particular, an implementation of the CM in a general population survey would be a sensible next step for evaluating the feasibility of the CM. Although the CM is less complicated than the RRT and seemed to work well with our university students, it might still be too cumbersome for other populations.⁹

Second, in the present study, no external validation data were available to evaluate the accuracy of the respondents' answers. Because plagiarism is socially undesirable and can be severely sanctioned, we argue that such behavior is likely to be underreported in direct questioning and, hence, the "more-is-better" assumption is theoretically plausible. However, a better approach to evaluate whether the CM yields more accurate results than direct questioning and, if so, to what extent it reduces the social-desirability bias would be to compare the respondents' self-reports against some objective data containing "true scores" (e.g., using criminal record checks or using an experimental setting in

9. And, as a reviewer of this article pointed out, for certain research questions (e.g., drug abuse), the most interesting population groups might be the ones with the lowest tolerance for the additional effort required by the CM.

which respondents are given the opportunity to engage in the sensitive behavior). Therefore, we strongly encourage validation studies for the CM.

Third, similar to other techniques for sensitive questions, a limitation of the CM is that larger samples are required to achieve the same level of statistical precision compared to direct questioning. That is, there is a trade-off between bias and efficiency. The loss in statistical efficiency and the costs of doubling or tripling the sample size can be justified only if the amount of bias reduction by the CM is large. Therefore, the CM should be recommended only for situations in which the social-desirability bias is likely to be severe. Moreover, as is true for most RRT implementations, the efficiency loss of the CM is particularly pronounced for items with a very low (or very high) prevalence (as our results for severe plagiarism illustrate). That is, compared to direct measurement, the relative sampling variance of the CM estimator increases the closer the true prevalence gets to zero (or one). Unfortunately, sensitive behaviors of interest are often behaviors with low prevalence. Choosing a non-sensitive item in the CM that has a low prevalence itself can counterbalance the efficiency loss, but this might also reduce the level of privacy protection perceived by the respondents. This is, in fact, a general problem of sensitive-question techniques that induce anonymity by adding noise to the data (as does the CM or RRT): Low-prevalence questions tolerate little noise for the estimates to remain efficient. How different levels of noise affect the respondents' trust in the techniques is, however, an issue for future research.

Fourth, it has to be mentioned that the statistical analysis of data gathered by the CM is more difficult than the analysis of data from direct questions because standard techniques cannot be employed without modification. We presented formulas for one important class of applications in this article (regression models). How to implement other statistical techniques with CM data has yet to be investigated, although solutions to similar problems for RRT data might prove to be helpful.

Appendix A. Implementation of the Crosswise Model in Our Study (translated from German)

In the following part of the questionnaire we apply a novel questioning technique to provide additional protection of your privacy for questions that might appear unpleasant or embarrassing to you. Please carefully read the instructions before answering the questions.

Two questions will be asked together at a time in blocks. Please first think about how you would answer each question separately (either with *Yes* or with *No*) but do not write these answers down. After that, please mark option (A) or option (B) depending on your answers according to the following rules:

If your answer is *No* to both questions or *Yes* to both questions, tick option (A)

If your answer is *Yes* to one of the questions and *No* to the other, tick option (B).

Your privacy remains protected because we do not know your answers to the separate questions. With the help of statistical procedures, however, we can compute to how many people overall the second question applies.

Block 1

Question 1: *Is your mother's birthday in January, February or March?*

Question 2: *When writing an assignment (e.g. seminar paper, term paper, thesis), have you ever intentionally adopted a passage from someone else's work without citing the original?*

How are your answers to the two questions?

- ☐ (A) **No to both questions or Yes to both questions**
☐ (B) **Yes to one of the questions and No to the other**

Block 2

Question 1: *Is your father's birthday in October, November or December?*

Question 2: *Did you ever have someone else write a large part of an assignment for you or hand in someone else's work (e.g. from www.hausarbeiten.de) as your own?*

How are your answers to the two questions?

- ☐ (A) **No to both questions or Yes to both questions**
☐ (B) **Yes to one of the questions and No to the other**

References

- Boruch, Robert F. 1971. "Assuring Confidentiality of Responses in Social Research: A Note on Strategies." *American Sociologist* 6(4):308–11.
- Coutts, Elisabeth, and Ben Jann. 2011. "Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT)." *Sociological Methods & Research* 40(1):169–93.
- Cruyff, Maarten J. L. F., Ardo van den Hout, Peter G. M. van der Heijden, and Ulf Böckenholt. 2007. "Log-Linear Randomized-Response Models Taking Self-Protective Response Behavior into Account." *Sociological Methods and Research* 36(2):266–82.
- Culwin, Fintan. 2006. "An Active Introduction to Academic Misconduct and the Measured Demographics of Misconduct." *Assessment & Evaluation in Higher Education* 31(2):167–82.
- ETH Zurich Executive Board. 2004. "Disciplinary Code of the Swiss Federal Institute of Technology Zurich (ETH Zurich Disciplinary Code)." Available at http://www.rechtssammlung.ethz.ch/pdf/361.1_Disziplinarordnung_engl.pdf.
- Fox, James Alan, and Paul E. Tracy. 1986. *Randomized Response: A Method for Sensitive Surveys*. Newbury Park, CA: Sage.
- Holbrook, Allyson L., and Jon A. Krosnick. 2010. "Social-Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." *Public Opinion Quarterly* 74(1):37–67.
- Jann, Ben. 2005. "rlogit: Stata Module to Estimate Logistic Regression for Randomized Response Data." Available at <http://ideas.repec.org/c/boc/bocode/s456203.html>.
- Krohn, Wolfgang, Corinna Schlombs, and Niels-Christian Taubert. 2003. "Plagierte Hausarbeiten: Problemlage an der Universität Bielefeld." Available at <http://www.uni-bielefeld.de/Benutzer/MitarbeiterInnen/Plagiate/iug2001.html>.
- Krumpal, Ivar. 2008. "Evaluation of the Effectiveness of the Randomized Response Technique and the Item Count Method in the Telephone Survey Mode." In *Proceedings of the 7th International Conference on Social Science Methodology, RC33—Logic and Methodology in Sociology (ISA)*, CD-ROM, eds. Simona Balbi, Germana Scepi, Giorgio Russolillo, and Agnieszka Stawinoga, Naples: Italy Jovene Editore.
- Landsheer, Johannes A., Peter van der Heijden, and Ger van Gils. 1999. "Trust and Understanding, Two Psychological Aspects of Randomized Response: A Study of a Method for Improving the Estimate of Social Security Fraud." *Quality & Quantity* 33(1):1–12.
- Lensvelt-Mulders, Gerty J. L. M., and Hennie R. Boeije. 2007. "Evaluating Compliance with a Computer-Assisted Randomized-Response Technique: A Qualitative Study into the Origins of Lying and Cheating." *Computers in Human Behavior* 23(1):591–608.
- Lensvelt-Mulders, Gerty J. L. M., Joop J. Hox, Peter G. M. van der Heijden, and Cora J. M. Maas. 2005. "Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation." *Sociological Methods and Research* 33(3):319–48.
- Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.
- Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge, UK: Cambridge University Press.
- Marsden, Helen, Marie Carroll, and James T. Neill. 2005. "Who Cheats at University? A Self-Report Study of Dishonest Academic Behaviors in a Sample of Australian University Students." *Australian Journal of Psychology* 57(1):1–10.
- Mood, Carina. 2010. "Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It." *European Sociological Review* 26(1):67–82.
- Office of Science and Technology Policy. 2000. "Executive Office of the President; Federal Policy on Research Misconduct; Preamble for Research Misconduct Policy." *Federal Register*, 65(235):76,260–64.

- Ostapczuk, Martin, Jochen Musch, and Morten Moshagen. 2009. "A Randomized-Response Investigation of the Education Effect in Attitudes Toward Foreigners." *European Journal of Social Psychology* 39(6):920–31.
- Park, Chris. 2003. "In Other (People's) Words: Plagiarism by University Students—Literature and Lessons." *Assessment & Evaluation in Higher Education* 28(5):471–88.
- Sattler, Sebastian. 2007. *Plagiate in Hausarbeiten: Erklärungsmodelle mit Hilfe der Rational Choice Theorie*. Hamburg, Germany: Verlag Dr. Kovač.
- Scheers, N. J., and C. Mitchell Dayton. 1988. "Covariate Randomized Response Models." *Journal of the American Statistical Association* 83(404):969–74.
- Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133(5):859–83.
- van der Heijden, Peter G. M., Ger van Gils, Jan Bouts, and Joop J. Hox. 2000. "A Comparison of Randomized Response, Computer-Assisted Self-Interview, and Face-to-Face Direct Questioning: Eliciting Sensitive Information in the Context of Welfare and Unemployment Benefit." *Sociological Methods and Research* 28(4):505–37.
- Warner, Stanley L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60(309):63–69.
- . 1986. "The Omitted Digit Randomized-Response Model for Telephone Applications." In *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 441–43. Available at: <http://www.amstat.org/sections/srms/proceedings/>.
- Yu, Jun-Wu, Guo-Liang Tian, and Man-Lai Tang. 2008. "Two New Models for Survey Sampling with Sensitive Characteristic: Design and Analysis." *Metrika* 67(3):251–63.