

On the comprehensibility and perceived privacy protection of indirect questioning techniques

Adrian Hoffmann\*<sup>1</sup>, Berenike Waubert de Puiseau\*<sup>1</sup>, Alexander F. Schmidt<sup>2</sup>, Jochen Musch<sup>1</sup>

<sup>1</sup>University of Duesseldorf, <sup>2</sup>University of Luxembourg

«fn»\*A. Hoffmann and B. Waubert de Puiseau contributed equally to this work.

#### Author Note

<sup>1</sup> Adrian Hoffmann, Berenike Waubert de Puiseau & Jochen Musch, Department of Experimental Psychology, University of Duesseldorf.

<sup>2</sup> Alexander F. Schmidt, Institute for Health and Behavior, Integrative Research Unit on Social and Individual Development, University of Luxembourg.

Correspondence concerning this article should be addressed to Adrian Hoffmann, Department of Experimental Psychology, University of Duesseldorf, Universitaetsstrasse 1, Building 23.03, 40225 Duesseldorf, Germany. E-mail: adrian.hoffmann@hhu.de

### Abstract

On surveys that assess sensitive personal attributes, indirect questioning aims at increasing respondents' willingness to answer truthfully by protecting confidentiality. However, the assumption that subjects understand questioning procedures fully and trust them to protect their privacy is tested rarely. In a scenario-based design, we compared four indirect questioning procedures in terms of comprehensibility and perceived privacy protection. All indirect questioning techniques were found less comprehensible for respondents than a conventional direct question used for comparison. Less-educated respondents experienced more difficulties when confronted with any indirect questioning technique. Regardless of education, the Crosswise Model was found most comprehensible among the four indirect methods. Indirect questioning was perceived to increase privacy protection in comparison to a direct question. Unexpectedly, comprehension and perceived privacy protection did not correlate. We recommend assessing these factors separately in future evaluations of indirect questioning.

*Keywords:* confidentiality, comprehension, randomized response technique, stochastic lie detector, crosswise model

On the comprehensibility and perceived privacy protection of indirect questioning techniques

When queried about sensitive personal attributes, some respondents conceal their true statuses by responding untruthfully to present themselves in a socially desirable manner (Krumpal, 2013; Marquis, Marquis, & Polich, 1986; Tourangeau & Yan, 2007). To increase respondents' willingness to respond honestly, indirect questioning procedures such as the randomized response technique (Warner, 1965) enhance the confidentiality of individual answers to sensitive questions. Consequently, prevalence estimates for sensitive personal attributes obtained through indirect questioning are considered more valid than prevalence estimates based on conventional direct questioning. However, use of indirect questioning relies on the assumption that participants understand all instructions, and understand how the procedures increase privacy protection (Landsheer, van der Heijden, & van Gils, 1999). Violation of this assumption is potentially at odds with a method's acceptance and validity of results. Employing a quasi-experimental design, this study investigates the influence of questioning techniques and education on comprehension and perceived privacy protection. Four indirect questioning techniques were investigated, and a conventional direct question served as a control condition.

### **Indirect Questioning Techniques**

To minimize bias due to respondents not answering truthfully to a sensitive question, Warner (1965) introduced the randomized response technique (RRT). With the original RRT procedure, respondents are confronted simultaneously with two related questions: a sensitive question A ("Do you carry the sensitive attribute?") and its negation question B ("Do you not carry the sensitive attribute?"). Participants answer one of these two questions, depending on the outcome of a randomization procedure, which is known only to the respondent and not the experimenter. When using a die as a randomization device, for example, respondents might be

asked to answer question A if the die shows a number between 1 and 4 (randomization probability  $p = 4/6$ ), and to answer question B if the die shows either 5 or 6 ( $p = 2/6$ ). Hence, a “Yes” response does not allow conclusions regarding a respondent’s true status. He or she might be a carrier of the sensitive attribute who was instructed to respond to statement A, or a non-carrier instructed to respond to B. Since the randomization probability  $p$  is known, the proportion of carriers of the sensitive attribute  $\pi$  can be estimated at the sample level (Warner, 1965). Since the collection of individual data related directly to the sensitive attribute is avoided, respondents queried about sensitive topics are expected to answer more truthfully when asked indirectly, rather than through direct questioning (DQ). Prevalence estimates obtained via RRT are supposed to exceed DQ estimates, and this has been found repeatedly (Lensvelt-Mulders, Hox, van der Heijden, & Maas, 2005). However, non-significantly different estimates in RRT and DQ conditions, and estimates higher in the DQ than in the RRT condition, have also been reported (e.g., Holbrook & Krosnick, 2010; Wolter & Preisendörfer, 2013). Moreover, given identical sample sizes, RRT estimates are always accompanied by a higher standard error than DQ since employing randomization adds unsystematic variance to the estimator (Ulrich, Schröter, Striegel, & Simon, 2012).

Following the original model from Warner (1965), various, more advanced RRT models have been proposed that focus on optimizing the statistical efficiency, validity, and applicability of the method (e.g., Dawes & Moore, 1980; Horvitz, Shah, & Simmons, 1967; Mangat & Singh, 1990). Several reviews and monographs provide detailed descriptions of RRT models and their applications (e.g., Chaudhuri & Christofides, 2013; Fox & Tracy, 1986; Umesh & Peterson, 1991). We present four indirect questioning procedures used in studies that investigate the

prevalence of sensitive, personal attributes, and compare them in terms of comprehensibility and perceived privacy protection.

### **The Cheating Detection Model**

With the cheating detection model (CDM; Clark & Desharnais, 1998), participants are confronted with a forced-response paradigm. After presentation of a single, sensitive question, the outcome of a randomization procedure determines whether respondents answer truthfully to this question with probability  $p$  or ignore the question and answer “Yes” with probability  $1-p$ . Since the outcome of the randomization procedure remains confidential, a “Yes” response does not allow for conclusion concerning an individual’s status with respect to a sensitive attribute. Clark and Desharnais (1998) suspect some participants disobey instructions by responding “No” regardless of the outcome of randomization, to avoid risk of being marked as a carrier of a sensitive attribute. Consequently, three disjoint and exhaustive classes are considered with CDM: carriers of the sensitive attribute responding truthfully ( $\pi$ ), honest non-carriers ( $\beta$ ), and respondents concealing their true statuses by answering “No” without regard for instructions. Clark and Desharnais refer to the latter class as cheaters ( $\gamma$ ). An example of a CDM question using a respondent’s month of birth as a randomization device is shown in Figure 1.

#### *TAKE IN FIGURE 1*

The CDM has been shown repeatedly to produce higher, and thus presumably more valid, prevalence estimates than direct questions or other indirect questioning techniques that do not consider instruction disobedience (e.g., Ostapczuk, Musch, & Moshagen, 2011). Validation studies arrive frequently at estimates of  $\gamma$  that exceed zero substantially, demonstrating the usefulness of a cheating-detection approach (e.g., Moshagen, Musch, Ostapczuk, & Zhao, 2010). However, in the case of  $\gamma > 0$ , the CDM provides only a lower and upper bound for the

proportion of carriers since the true statuses of respondents classified as cheaters are unknown. Hence, the rate of carriers could be located within the range of  $\pi$  (were no cheater a carrier) and  $\pi + \gamma$  (were all cheaters carriers).

### **The Stochastic Lie Detector**

Similar to the original RRT procedure (Warner, 1965), the recently proposed stochastic lie detector (SLD; Moshagen, Musch, & Erdfelder, 2012) confronts respondents with sensitive question A and its negation B. Similar to the modified RRT model that Mangat (1994) proposes, only part of the participants is instructed to engage in randomization. Carriers of the sensitive attribute respond to question A unconditionally, and if they respond truthfully, their answer should always be “Yes”. Non-carriers respond to question A with randomization probability  $p$ , and to question B with probability  $1-p$ . Consequently, neither a “Yes” nor “No” response unequivocally reveals a respondent’s true status. However, Moshagen et al. (2012) argue that some carriers of the sensitive attribute might feel a desire to lie and respond “No”, even if instructed otherwise. This assumption was represented by a new parameter  $t$ , which accounts for the proportion of carriers answering truthfully, while the remaining proportion of the carriers ( $1-t$ ) are assumed to lie about their statuses. In contrast, non-carriers should not have any reason to lie. An example of an SLD question is shown in Figure 2.

#### *TAKE IN FIGURE 2*

During a pilot study, application of the SLD resulted in a prevalence estimate for domestic violence that exceeded an estimate obtained using a direct question. Moreover, the SLD estimated the proportion of nonvoters in the German federal elections in 2009 in concordance with the known true prevalence (Moshagen et al., 2012). In a second study by Moshagen, Hilbig, Erdfelder, and Moritz (2014), cheating behaviors were induced experimentally to allow direct

determination of the proportion of cheaters as an external validation criterion. Again, SLD closely reproduced the known proportion of carriers of the sensitive attribute, while DQ produced an underestimate. In contrast to these results, a recent experimental comparison of SLD with competing questioning techniques found SLD to overestimate the known prevalence of a non-sensitive control question (Hoffmann & Musch, 2015). Although this mixed pattern of results might be explained in terms of sampling error, difficulties regarding understanding SLD instructions offer an alternative explanation.

### **The Crosswise Model**

A new class of non-randomized response techniques was proposed recently (Tian & Tang, 2014), offering simplified assessment of the prevalence of sensitive attributes since no external randomization device is required. One of the most promising candidates among these is the crosswise model (CWM; Yu, Tian, & Tang, 2008) because it offers symmetric answer categories (i.e., none of the answer options is a safe alternative that eliminates identification as a carrier). With CWM, participants are presented with two statements simultaneously: one statement refers to the sensitive attribute with unknown prevalence  $\pi$ , and a second to a non-sensitive control attribute with known prevalence  $p$  (e.g., a respondent's month of birth). Participants indicate whether "both statements are true or both statements are false", or whether "exactly one of the two statements is true (irrespective of which one)". If an individual respondent's month of birth is unknown to the questioner, CWM grants confidentiality of respondents' true statuses, presumably leading to undistorted prevalence estimates for sensitive attributes. Figure 3 shows an example of a CWM question.

*TAKE IN FIGURE 3*

In various studies, application of CWM resulted in higher prevalence estimates for sensitive attributes than DQ (e.g., Coutts, Jann, Krumpal, & Näher, 2011; Kundt, Misch, & Nerré, 2013). An experimental comparison of CWM, SLD, and a DQ condition showed that CWM and SLD prevalence estimates of xenophobia and Islamophobia exceeded those obtained via DQ (Hoffmann & Musch, 2015). In another study, the CWM estimated the known prevalence of experimentally induced cheating behavior accurately (Hoffmann, Diedenhofen, Verschuere, & Musch, 2015). Yu et al. (2008) argue that non-randomized models are “easy to operate for both interviewer and interviewee” (p. 261), which offers an explanation for promising results observed to date using the CWM.

### **The Unmatched Count Technique**

Introduced by Miller (1984), the unmatched count technique (UCT) also offers comparably simple instructions. Respondents are assigned randomly to an experimental or control group, both of which are confronted with a list of non-sensitive statements. In the experimental group, the list additionally contains a sensitive statement. In both groups, respondents indicate how many, but not which, of the statements apply to them. Since the only disparity between the two groups is the addition of a question referring to the sensitive attribute in the experimental group, a difference in mean reported total counts estimates the proportion  $\pi$  of carriers of the sensitive attribute (Erdfelder & Musch, 2006; Miller, 1984). The individual statuses of respondents in the experimental group remain confidential as long as the total reported count is different from zero (in which case all statements could be deduced to have been answered negatively), and different from the maximum count possible (in which case all statements, including the sensitive statement, could be deduced to have been answered affirmatively). Thus, experimenters should prevent such extreme counts cautiously by including

a sufficient number of non-sensitive statements (Erdfelder & Musch, 2006; Fox & Tracy, 1986).

An example of a UCT question with one sensitive and three non-sensitive items is shown in Figure 4.

*TAKE IN FIGURE 4*

UCT has repeatedly provided higher prevalence estimates for sensitive attributes than DQ approaches did (e.g., Ahart & Sackett, 2004; Coutts & Jann, 2011; Wimbush & Dalton, 1997). Comprehensibility of the instructions and trust in the method were found to exceed that of the RRT and a conventional DQ approach (Coutts & Jann, 2011). These results however were limited to a comparison of UCT and a forced-response RRT design, and comprehension was evaluated only by means of potentially forgeable self-ratings.

A meta-analytic evaluation of indirect questioning studies (Lensvelt-Mulders et al., 2005) reveals that prevalence estimates obtained through RRT largely meet the more-is-better criterion; i.e., RRT estimates for socially undesirable attributes exceeding estimates based on DQ indicate increased validity since social desirability biases them less. Another meta-analytic accumulation of strong validation studies in which the known true prevalence of a sensitive attribute served as an objective criterion found that RRT yields prevalence estimates that are substantially less biased than DQ estimates (Lensvelt-Mulders et al., 2005). Some studies present RRT estimates that are indifferent from (e.g., Kulka, Weeks, & Folsom, 1981) or even lower than (e.g., Holbrook & Krosnick, 2010) DQ estimates. Regarding thorough examination of the validity of indirect questioning, in some strong validation studies, RRT estimates deviated substantially from known population values (e.g., Kulka et al., 1981; van der Heijden, van Gils, Bouts, & Hox, 2000). These results might be explained in terms of participants' noncompliance with instructions even under RRT conditions, especially concerning surveys that cover highly

sensitive personal attributes (e.g., Clark & Desharnais, 1998; Edgell, Himmelfarb, & Duchan, 1982; Moshagen et al., 2012). Two psychological aspects that are likely to play a role in respondents' willingness to cooperate are a) the ability to understand instructions and b) whether respondents trust the promise of confidentiality associated with use of indirect questioning.

### **Comprehensibility and perceived privacy protection from indirect questioning**

Most indirect questioning relies on the assumption that participants comply with instructions; they are capable and willing to cooperate (Abul-Ela, Greenberg, & Horvitz, 1967; Edgell et al., 1982). Many researchers raise concerns that some participants might not understand instructions for indirect questions fully since they are generally more complex in comparison to DQ (Coutts & Jann, 2011; Landsheer et al., 1999). Participants might also not trust indirect questioning to protect their privacy, and might therefore disregard instructions (Clark & Desharnais, 1998; Landsheer et al., 1999). Response bias resulting from lack of understanding or trust toward a method threatens the validity of prevalence estimates determined through indirect questions (Holbrook & Krosnick, 2010; James, Nepusz, Naughton, & Petroczi, 2013). Hence, trust and understanding are two psychological factors that determine the validity of indirect questioning (Fox & Tracy, 1980; Landsheer et al., 1999).

One strategy used to evaluate the comprehensibility and perceived privacy protection is assessment of response rates in surveys that use indirect questioning. Following the logic of these studies, higher response rates indicate higher trust and understanding. While some studies show reduced response rates in RRT conditions compared to DQ (Coutts & Jann, 2011), other studies report comparable response rates for indirect and direct questioning (e.g., I-Cheng, Chow, & Rider, 1972; Locander, Sudman, & Bradburn, 1976), or higher response rates during indirect questioning (e.g., Fidler & Kleinknecht, 1977; Goodstadt & Gruson, 1975). However, these

results only allow indirect conclusions regarding the comprehensibility and perceived privacy protection of the questioning techniques used since there exist numerous alternative explanations for disparities in response rates (e.g., motivational factors and the content of sensitive questions). Therefore, differential influences of trust and understanding cannot be disentangled based on analysis of response rates.

Using more controlled approaches, some validation studies use known individual statuses of respondents regarding sensitive attributes to determine whether they responded in accordance with instructions. The rate of demonstrably untrue responses was used to estimate the rate of participants who did not understand or trust the questioning procedure. Edgell et al. (1982) and Edgell, Duchan, and Himmelfarb (1992) argue that low rates of 2% to 4% incorrect responses to moderately sensitive questions indicate a high level of comprehension. However, the rate of false answers rose to 10% to 26% for highly sensitive questions. It is plausible that this stronger bias might in part be caused by respondents distorting answers to increasingly distance themselves from more sensitive attributes (Edgell et al., 1982). A meta-analytic investigation of strong validation studies in which participants' true statuses concerning a sensitive attribute was known identified a mean rate of 38% incorrect responses for RRT questions, while other questioning formats produced up to 49% false answers (Lensvelt-Mulders et al., 2005). Disparities between RRT and DQ estimates increased for questions with higher sensitivity. This pattern could be interpreted as evidence that respondents trust the confidentiality offered by indirect questioning but require enhanced privacy protection, and use it only if a sensitive issue is at stake. However, designs used in these studies did not separate the influences of comprehension and perceived privacy protection.

A more direct strategy to determine trust and understanding for varying questioning procedures is to assess these two constructs directly on a survey. Various studies based on reports of interviewees and interviewers estimated the rate of respondents fully understanding the RRT procedure at 94% (I-Cheng et al., 1972), 78% to 90% (Locander et al., 1976), 79% to 83% (van der Heijden, van Gils, Bouts, & Hox, 1998) and 80% to 93% (Coutts & Jann, 2011). For the UCT (Miller, 1984), the rate was 92%. In another study, the comprehensibility of an RRT question was rated as normal or easy by 89% of respondents, and 10% indicated it was difficult (Hejri, Zendehdel, Asghari, Fotouhi, & Rashidian, 2013).

To estimate trust toward an RRT question, some researchers asked participants whether they thought there was a trick to the RRT procedure. Since 20% to 40% (Abernathy, Greenberg, & Horvitz, 1970) and 15% to 37% (I-Cheng et al., 1972) of respondents answered affirmatively to this statement, a considerable fraction of respondents appear to mistrust RRT despite a promise of confidentiality. When confronted with an indirect question, respondents estimated the probability of the researcher knowing which questions they answered at 55% to 72% (Soeken & Macready, 1982). Consequently, the probability of the procedure granting confidentiality was estimated at only 28 to 45%. Few respondents (15% to 22%) believed that RRT guaranteed the anonymity of their answers in a study from Coutts and Jann (2011); for a UCT question, the rate was slightly higher though low at 29%.

Aside from assessment of total rates of trust and understanding, some studies compare perceived privacy protection of direct versus indirect questions. In one study, 91% of respondents felt that the RRT would enhance confidentiality compared to DQ (Edgell et al., 1982). In another, a rate of 72% of respondents trusting the RRT procedure was unexpectedly exceeded by a rate of 83% of trustful participants in a DQ condition (van der Heijden et al.,

1998), implying RRT failed to establish higher trust. Only 29% of participants in a study from Hejri et al. (2013) perceived that the RRT increased confidentiality when compared to DQ. Other studies comparing indirect questioning techniques indicated that the UCT might be superior to RRT regarding trust and understanding (Coutts & Jann, 2011; James et al., 2013).

Few studies examined the influence of cognitive skill and education on comprehension and perceived privacy protection of indirect questioning designs. I-Cheng et al. (1972) found a positive effect of education on rate of cooperative respondents. While 72% of participants failed to understand an RRT question, the rate dropped to 27% for participants who graduated from primary school and to 2% for participants who held a junior high school degree. Landsheer et al. (1999) found no influence of participants' formal education on incidences of incorrect answers. Holbrook and Krosnick (2010) report that the most implausible results in their study occurred in a subgroup of highly educated participants, indicating that the "failure of the RRT was not due to the cognitive difficulty of the task" (p. 336).

Overall, results from studies that investigate participants' trust in and understanding of indirect questioning are inconclusive. Some studies report high rates of trust and understanding and others show that a substantial share of participants fail to understand indirect questions, or do not trust the procedures. Data do not allow separation of these factors, and thus independent assessment of trust and understanding is needed to identify indirect questioning techniques that are both comprehensible and trustful. The role of cognitive skill and education as moderators of trust and understanding is not yet understood.

### **Present Study**

In this study, four indirect questioning techniques used frequently in survey research that addresses sensitive questions were entered into an experimental comparison of comprehensibility

and perceived privacy protection. The CDM (Clark & Desharnais, 1998) and the SLD (Moshagen et al., 2012) allow for a separate estimation of the proportion of noncompliant respondents in the sample by implementing an additional cheating parameter. The CWM (Yu et al., 2008) is presumably easier to understand than other RRT models and offers a symmetric design, which might facilitate honest responding. The UCT (Miller, 1984) is similarly easy to employ, and some participants prefer UCT over RRT questions concerning trust and understanding. This study evaluates the comprehensibility and perceived privacy protection of these four indirect questioning techniques separately since these two factors might be intertwined though not linked causally in a unidirectional connection. Some participants might understand the instructions but not trust the protection of their privacy. Others might fail to comprehend the task but perceive that indirect questions offer more confidentiality than conventional direct questioning approaches do.

To allow an objective and rigorous evaluation of participants' instruction comprehension, we used a scenario-based design. To assess whether they understood the procedure, participants responded to a number of questions vicariously for various fictional characters. Participants were first given information regarding these characters (e.g., "Wilhelm has never cheated on an exam" or "Wilhelm was born on July"), were subsequently provided with instructions for one of the indirect questioning techniques, and finally indicated which answer the fictional character must give. This approach ensured participants would not respond untruthfully to conceal personal statuses regarding sensitive attributes. As a benefit of the scenario-based design, the true status for each fictional character was known to both the respondent and questioner, and thus served as an objective criterion for assessment of the correctness of a respondent's answers. The mean proportion of questions answered correctly in a test that assessed a respondent's understanding of

the procedure was determined as an estimate of the comprehensibility of each questioning procedure. We also assessed how participants estimated the privacy protection offered by various questioning techniques. Finally, by questioning two groups of participants with high versus low educations, we investigated moderation of cognitive skill.

This study addresses the following research questions: 1) Do indirect questions differ from conventional direct questions regarding comprehensibility? If so, which one of the four models under investigation is most comprehensible? 2) Do indirect questions offer higher perceived privacy protection than direct questions do? If so, what model is perceived as most protective? 3) Do cognitive skills, measured by respondents' education, moderate the influence of questioning technique on comprehension or perceived privacy protection? 4) Is there an association between comprehension and perceived privacy protection?

## **Methods**

### **Participants**

Seven-hundred sixty-six participants were recruited to participate in an online survey through a commercial online panel. Since education was part of the experimental design, an online quota ensured matching proportions of participants with lower versus higher educations. Participants in the lower-education group finished at most nine years of school (the German *Hauptschule*), and participants in the higher-education group finished at least twelve years of education (the German *Abitur*). To optimize statistical power to detect differences between experimental conditions, we decided to increase homogeneity of our sample by allowing only respondents between 25 and 35 years of age to participate. This particular range was chosen because it matches the age range of the respondents that participate most often in online studies (Gosling, Vazire, Srivastava, & John, 2004). Of the initially invited participants, 171 (22%) were

rejected due to full quotas, 58 (8%) were screened out at the first page of the questionnaire as they did not match the inclusion criteria (education and age range), and 136 (18%) were excluded as they failed to complete the questionnaire. Of the 136 participants who started but did not complete the questionnaire, 41 (5% of the initially invited) aborted the experiment before any of the experimental questions were presented; 95 (12% of the initially invited) viewed at least one of the questioning techniques. To test for selective dropout with respect to experimental conditions, we compared which type of question participants saw last before dropping out ( $N = 95$ ). As a reference, we compare these proportions against those of the last type of question for participants completing the study ( $N = 401$ ). Within CDM (21% vs. 22%), CWM (23 vs. 21%), and UCT (18% vs. 20%) conditions, distributions did not differ between incomplete and complete data sets. There was a trend towards a lower dropout rate in the more simple DQ condition (6% vs. 16%), and a higher dropout rate in the more complex SLD condition (32% vs. 21%); this trend was however small and insignificant,  $\chi^2(4, N = 496) = 8.55, p = .07, w = .13$ . Educational level (high vs. low) did not differ between aborting and finishing participants either,  $\chi^2(1, N = 496) = 2.67, p = .10, w = .07$ . Participants in the final sample ( $N = 401$ , 52% of the initially invited) had a mean age of 30.72 years ( $SD = 3.35$ ); 211 (53%) were female and 386 (97%) indicated German as their first language. Education groups were represented evenly, with 199 lower- and 202 higher-education participants. Power analyses conducted using G\*Power 3 software (Faul, Erdfelder, Buchner, & Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007) revealed that our large sample size provided sufficient power for detection of medium effects during analysis of mean differences between groups ( $f = 0.25$ ;  $1-\beta = .99$ ) and (both parametric and nonparametric) correlations ( $r / r_S = .30$ ;  $1-\beta > .99$ ).

## Design

The scenario-based experiment implemented a 5 (questioning technique) by 2 (educational level), quasi-experimental mixed design. Questioning technique varied within subjects, realized in five blocks: CDM (Clark & Desharnais, 1998), SLD (Moshagen et al., 2012), CWM (Yu et al., 2008), UCT (Miller, 1984), and a conventional DQ approach. The second, quasi-experimental, between-subjects independent variable was the participants' education (high versus low).

Academic cheating served as the sensitive attribute, as used in several studies of indirect questioning techniques (e.g., Hejri et al., 2013; Lamb & Stem, 1978; Ostapczuk, Moshagen, Zhao, & Musch, 2009; Scheers & Dayton, 1987). The wording of the sensitive question was identical in all questioning technique conditions, reading "Have you ever cheated on an exam?" Three additional, non-sensitive attributes were used to employ indirect questioning techniques. First, month of birth was used as the randomization device for the CDM, SLD, and CWM questions. To allow application of the UCT format, we constructed a list of four items: the sensitive attribute, the non-sensitive month of birth, and two non-sensitive attributes (i.e., gender and a question concerning whether participants visited London). Indirect questioning techniques were implemented as shown in Figures 1 through 4. Each of the questioning techniques was applied to four fictional characters named Ludwig, Ernst, Hans, and Wilhelm, characterized differently regarding the sensitive and non-sensitive attributes. Ludwig and Ernst were presented as carriers of the sensitive attribute, and Hans and Wilhelm were described as non-carriers. The birthdays of Ludwig and Hans were chosen to fall into one of the outcome categories of the binary randomization procedure, and the months of birth for Ernst and Wilhelm were set to fall into the other category. All four characters were male, and none was described to have visited

London. The descriptions were chosen to avoid extreme counts in the UCT condition.

Descriptions of the four fictional characters were accessible to participants at any time during the experiment. To control for effects of serial position, the sequence of presentation of the five questioning technique blocks was randomized among participants. Additionally, the four fictional characters were presented in random order within each of the questioning technique blocks.

To examine the comprehensibility of the questioning techniques, participants vicariously indicated answers that the four fictional characters must give if confronted with each of the various questioning techniques. Descriptions of the characters were displayed along with the questions. As an example, a screenshot of a CWM question that had to be answered from the perspective of Wilhelm is shown in Figure 5. The comprehensibility of the questioning techniques was operationalized as the percentage of correct answers computed across all four fictional characters, separately for each participant.

#### *TAKE IN FIGURE 5*

To assess perceived privacy protection, participants rated perceived confidentiality offered by each questioning technique on a 7-point, Likert-type scale, ranging from -3 (no confidentiality) to +3 (perfect confidentiality). Scales were presented directly below the comprehension questions. Perceived privacy protection was operationalized as the mean score on these Likert-scales concerning all four fictional characters.

## **Results**

### **Comprehensibility**

Mean proportions of correct responses as a function of questioning technique and education are shown in Figures 6 and 7, respectively. Reliability analyses for the proportion of

correct responses across all five questioning techniques revealed that the variable measured a homogenous construct (Cronbach's  $\alpha = .75$ ). Descriptively, the mean proportion of correct responses in the DQ control condition was higher than with CDM ( $\Delta M = 15.04\%$ ,  $r = .44$ ,  $d_z = 0.70$ ; according to Cohen, 1988), SLD ( $\Delta M = 21.73\%$ ,  $r = .23$ ,  $d_z = 0.79$ ), CWM ( $\Delta M = 7.07\%$ ,  $r = .49$ ,  $d_z = 0.33$ ), and UCT ( $\Delta M = 13.38\%$ ,  $r = .52$ ,  $d_z = 0.49$ ) condition. Among indirect questioning techniques, mean proportion of correct responses was descriptively highest in the CWM condition, followed by scores in UCT (CWM versus UCT:  $\Delta M = 6.3\%$ ,  $r = .52$ ,  $d_z = 0.23$ ), CDM (CWM versus CDM:  $\Delta M = 8.0\%$ ,  $r = .39$ ,  $d_z = 0.33$ ; UCT versus CDM:  $\Delta M = 1.7\%$ ,  $r = .42$ ,  $d_z = 0.06$ ), and SLD (CWM versus SLD:  $\Delta M = 14.7\%$ ,  $r = .29$ ,  $d_z = 0.52$ ; UCT versus SLD:  $\Delta M = 8.4\%$ ,  $r = .25$ ,  $d_z = 0.24$ ; CDM versus SLD:  $\Delta M = 6.7\%$ ,  $r = .38$ ,  $d_z = 0.26$ ) condition. Descriptive differences in the mean proportion of correct responses between participants with high versus low education were negligible in the DQ control condition ( $\Delta M = 1.39\%$ ,  $d = 0.07$ ). Within the CDM condition, people with lower education had slightly lower scores ( $\Delta M = 4.98\%$ ,  $d = 0.24$ ). For SLD ( $\Delta M = 9.70$ ,  $d = 0.41$ ), CWM ( $\Delta M = 7.61\%$ ,  $d = 0.34$ ), and UCT ( $\Delta M = 11.07\%$ ,  $d = 0.36$ ) conditions, lower education resulted in substantially lower mean proportion of correct responses. Considering the binary nature of correct/incorrect responses, inferential statistics were determined by establishing a generalized linear mixed model with a logit link function, implementing the fixed factors questioning technique (within-subjects), education (between-subjects), and the interaction of these two factors (cf. Jaeger, 2008). Responses were coded as incorrect (0; reference category) versus correct (1) and served as the criterion. A by-subject random intercept accounted for the dependency of measurements. This model revealed a significant main effect for within-subjects questioning technique ( $F(4, 8010) = 77.51$ ,  $p < .001$ ). Sequentially Bonferroni-corrected pairwise contrasts for within-

subjects questioning technique widely mirrored descriptive results: comprehensibility in the DQ control condition was higher than with CDM ( $t(8010) = -5.64, p < .001$ ), SLD ( $t(8010) = -10.41, p < .001$ ), CWM ( $t(8010) = -5.99, p < .001$ ), and UCT ( $t(8010) = -11.11, p < .001$ ) condition. Pairwise comparisons among indirect questioning techniques resulted in significant differences for all combinations (CDM versus SLD:  $t(8010) = -7.53, p < .001$ ; CDM versus UCT:  $t(8010) = -6.96, p < .001$ ; SLD versus CWM:  $t(8010) = 7.51, p < .001$ ; SLD versus UCT:  $t(8010) = 2.36, p < .05$ ; CWM versus UCT:  $t(8010) = -6.96, p < .001$ ), except for the difference between CDM and CWM that was not statistically reliable ( $t(8010) = -0.158, p = .88$ ). Thus, participants demonstrated highest comprehension for direct questions. Comprehension was slightly but significantly reduced for CWM and CDM questions. For CDM, comprehensibility was descriptively, but not significantly lower than for CWM. For UCT, comprehension was significantly reduced further; but it was still significantly higher than for SLD questions, for which comprehension was lowest. Furthermore, the established model revealed a significant main effect for between-subjects education ( $F(1, 8010) = 9.07, p < .01$ ). As hypothesized, higher education resulted in a higher proportion of correct responses. Finally, the model showed a significant interaction of the two factors questioning technique and education ( $F(4, 8010) = 5.58, p < .001$ ). Sequentially Bonferroni-corrected pairwise contrasts indicated that high versus low education did not result in significantly different proportions of correct responses in the DQ ( $t(8010) = -0.98, p = .33$ ) or CDM ( $t(8010) = -0.63, p = .53$ ) conditions, respectively. For SLD ( $t(8010) = -2.17, p < .05$ ), CWM ( $t(8010) = -3.36, p < .01$ ), and UCT ( $t(8010) = -4.65, p < .001$ ) conditions, lower education resulted in lower comprehension. Hence, although the proportions of correct responses were comparable between educational groups for DQ, education moderated comprehension in three of four indirect questioning formats.

TAKE IN FIGURE 6

TAKE IN FIGURE 7

### Perceived privacy protection

Mean ratings of perceived privacy protection as a function of questioning technique and education are shown in Figures 8 and 9, respectively. Reliability analyses for mean ratings of perceived privacy protection across all five questioning techniques revealed that the variable measured a homogenous construct ( $\alpha = .87$ ). A univariate 5 (questioning technique) by 2 (education), mixed-model ANOVA revealed a main effect for within-subjects questioning technique ( $F(4,1596) = 18.76, p < .001, \eta^2 = .05$ ), but no effect for between-subjects education ( $F(1,399) < 1$ ). However, the two factors showed an interaction ( $F(4,1596) = 9.21, p < .001, \eta^2 = .02$ ). A Bonferroni *post-hoc* test of the factor questioning technique revealed that mean scores in the DQ control condition were lower than with CDM ( $\Delta M = 0.26, p < .001; r = .57, dz = 0.19$ ), SLD ( $\Delta M = 0.25, p < .01; r = .53, dz = 0.18$ ), CWM ( $\Delta M = 0.39, p < .001; r = .39, dz = 0.25$ ), and UCT ( $\Delta M = 0.52, p < .001; r = .40, dz = 0.33$ ) conditions. *Post-hoc* tests between the indirect questioning techniques showed that the UCT format resulted in the highest scores, indifferent from scores in the CWM condition ( $\Delta M = 0.13, p = .21; r = .64, dz = 0.12$ ) but higher than scores with CDM ( $\Delta M = 0.26, p < .001; r = .61, dz = 0.22$ ) and SLD ( $\Delta M = 0.27, p < .001; r = .64, dz = 0.24$ ) conditions. Mean scores in the CWM condition were comparable to scores in the CDM ( $\Delta M = 0.13, p = .31; r = .61, dz = 0.11$ ) and SLD ( $\Delta M = 0.14, p = .10; r = .67, dz = 0.13$ ) conditions. Finally, CDM and SLD scores showed no difference ( $\Delta M = 0.01, p > .99; r = .65, dz = 0.01$ ). Combined, all indirect questioning techniques enhanced perceived privacy protection in comparison with a conventional DQ. Participants perceived the highest privacy protection when confronted with UCT and CWM questions, and perceived privacy ratings for

CWM, CDM, and SLD questions did not differ. Since no main effect of education emerged, results are only presented for the interaction of education and questioning technique. Five pairwise t-tests for independent groups on a Bonferroni-corrected  $\alpha$ -level (*corrected*  $\alpha = .05 / 5 = .01$ ) were computed to compare participants with high versus low education separately within each questioning technique condition. The comparisons revealed an education effect only in the DQ condition ( $\Delta M = 0.51$ ,  $t(399) = 3.35$ ,  $p < .001$ ;  $d = 0.33$ ), while education groups did not significantly differ on the corrected  $\alpha$  within CDM ( $\Delta M = 0.08$ ,  $t(399) = 0.64$ ,  $p = .53$ ;  $d = 0.07$ ), SLD ( $\Delta M = 0.10$ ,  $t(399) = 0.78$ ,  $p = .43$ ;  $d = 0.08$ ), CWM ( $\Delta M = 0.10$ ,  $t(399) = 0.77$ ,  $p = .44$ ;  $d = 0.07$ ), and UCT ( $\Delta M = 0.26$ ,  $t(399) = 1.98$ ,  $p = .05$ ;  $d = 0.20$ ) conditions. Hence, participants with lower education perceived higher privacy protection when confronted with a direct question than participants with higher education, and perceived privacy protection did not differ between education groups within indirect questioning conditions.

*TAKE IN FIGURE 8*

*TAKE IN FIGURE 9*

### **Association of comprehension and perceived privacy protection**

To investigate whether participants' comprehension of a questioning technique was associated with perceived privacy protection, bivariate Spearman-correlations were computed for the total sample, and separately for the two education groups (Table 1). Comprehension and perceived privacy protection showed no significant associations.

*TAKE IN TABLE 1*

## Discussion

In the present study, we compared four indirect questioning procedures in terms of comprehensibility and perceived privacy protection. A conventional direct question served as a control condition. Moderating effects of participants' level of education were investigated.

### Comprehensibility of indirect questioning techniques

All indirect questioning techniques showed lower comprehensibility in comparison to a DQ condition. Results accord with extant studies that suggest the instructions of indirect questions are more complex and thus more difficult to comprehend than direct questions (e.g., Böckenholt, Barlas, & van der Heijden, 2009; Coutts & Jann, 2011; Edgell et al., 1992; Landsheer et al., 1999; O'Brien, 1977). In a qualitative interview study, Boeije and Lensvelt-Mulders (2002) report that the reduced comprehensibility of indirect RRT questions might be explained partially by participants experiencing difficulties when “doing two things at the same time” (p. 30). Participants struggle to focus on RRT questions and the randomization procedure simultaneously. This experience applies to the present study since subjects had to integrate two types of information to identify the correct responses in all indirect questioning conditions: first the status of the fictional characters regarding a sensitive attribute, and second their statuses concerning non-sensitive randomization attribute(s). Results suggest that some indirect questioning formats showed better comprehensibility than others did; CWM appears to have been the most comprehensible format, corroborating Yu et al.'s (2008) assertion that CWM is easier to operate. Integrating two types of information or “doing two things at the same time” (Boeije & Lensvelt-Mulders, 2002, p. 30; also see Lensvelt-Mulders & Boeije, 2007, p. 598) might have been easiest for participants in the CWM condition since this questioning format incorporates the randomization procedure and the response to the sensitive statement in a single

step. Respondents simply have to read two answer options and identify the appropriate one. In contrast, comprehension was lowest in the SLD condition. A more detailed inspection of the SLD's instructions revealed that participants must make three sequential decisions to identify the correct response: a) decide whether the fictional character is a carrier of the sensitive attribute, b) identify the question that must be answered as determined by the randomization procedure (if the character is a non-carrier), and c) identify the correct response to the respective question.

Answering an SLD question therefore arguably is more difficult, and more prone to errors, than answering a CWM question. However, as this explanation is rather speculative, future studies should consider qualitative interviews similar to the one conducted by Boeije and Lensvelt-Mulders (2002) to shed further light on the exact mechanisms that account for differential comprehensibility of the four indirect questioning models investigated here.

The lower-education group demonstrated decreased comprehension of all indirect questioning techniques, with the exception of CDM. Researchers investigating the prevalence of sensitive personal attributes should consider that the comprehension of indirect questions might be reduced in samples that include less-educated participants, and should refrain from applying indirect questioning techniques if less-educated individuals report difficulties while completing a survey. This caveat should receive particular attention if education is expected to associate with the sensitive attribute under investigation (e.g., negative attitudes towards foreigners; cf. Ostapczuk, Musch, & Moshagen, 2009).

On the one hand, since a within-subjects scenario-based design was used, comprehension rates reported in this study are likely a lower boundary for the comprehensibility of questioning procedures under investigation. Mean comprehension in the DQ condition was high (> 90%) and unaffected by education, indicating participants were generally capable of answering questions

from the perspective of the four fictional characters. However, participants' comprehension would likely improve if they had to deal with only one questioning technique, and if they were not required to respond vicariously about fictional characters but for themselves. On the other hand, as remarked by one of the reviewers of this paper, participants in our study were provided with all relevant information on-screen, which possibly facilitated the identification of the correct response. In real applications, this information has to be retrieved from memory. Under applied conditions, issues with the retrieval of autobiographical information with respect to the sensitive and / or non-sensitive attributes may therefore make it more difficult to identify the correct response. Instructions for all indirect questioning procedures were kept as concise as possible. During real applications, more comprehensive instructions could be presented along with extended explanations, and can be combined with comprehension checks to ensure respondents understand the procedure. In contrast to many extant studies that use face-to-face questioning or paper-pencil tests, this study confronted participants with an online questionnaire that contained indirect questioning techniques. Although RRT yielded valid results in previous online studies (e.g., Musch, Bröder, & Klauer, 2001), a face-to-face setting offers better opportunities to assist participants who experience difficulties, and might help respondents achieve better comprehension and avoid errors when answering questions.

### **Perceived privacy protection**

Regarding perceived privacy protection, all indirect questioning techniques showed higher mean scores than a conventional DQ, suggesting participants developed higher trust toward indirect questions. The highest mean score was achieved in the UCT condition, followed by a slightly but insignificantly reduced mean score with CWM. Scores under CWM, CDM, and SLD were similar, though the latter two differed from the UCT condition. Education influenced

perceived privacy protection only in the DQ condition, with lower-education participants reporting higher perceived protection. This education effect did not occur in any indirect questioning condition. Hence, the influence of education on perceived privacy protection reduces to failure to understand that direct questions provide poorer privacy protection. When sensitive questions are assessed using indirect questioning, education might be negligible concerning perceived protection.

Comprehension did not associate with perceived privacy protection for the entire sample, or in the two education groups. This pattern suggests that although participants understood the instructions, they did not necessarily trust the procedure. Results also suggest respondents developed trust despite failure to comprehend instructions fully. Lack of association between comprehension and perceived privacy protection suggests the importance of examining differential impacts of these two constructs separately when assessing sensitive topics with indirect questioning techniques. To allow valid assessment of the prevalence of sensitive personal attributes, participants should ideally both understand and trust the questioning technique.

### **Limitations and future directions**

Several limitations to our study have to be acknowledged. For example, despite the successful separation of comprehension and perceived privacy protection, a confounding influence of task motivation on the comprehensibility of questioning techniques cannot be ruled out. While comprehension in the DQ condition was generally high, about 10% of the participants' responses were incorrect. This suggests that there was a potential lack of motivation among at least some participants. However, in a recent study, Baudson and Preckel (2015) found that in other rather simple cognitive tasks, the proportion of successful participants was also only

90%, and thus, close to the accuracy we observed in the DQ condition. This provides evidence for the notion that it is probably unrealistic to expect perfect scores in tasks like the ones we investigated.

Arguably, a lack of motivation is likely to exert a stronger influence on cognitively more demanding tasks, such as responding to indirect rather than direct questions. Our dropout analyses indeed showed a small (yet insignificant) trend indicating a lower dropout rate in the less cognitively demanding DQ condition, and a higher dropout rate in the presumably rather demanding SLD condition.

It is conceivable that participants with lower education might also be less motivated. However, given that comprehension in the DQ condition did not differ between high and low education groups, a general difference in motivation between these two groups seems to be rather unlikely. Moreover, while the design of our experiment did not allow us to directly observe evidence for a lack of motivation, any such motivational differences are likely to affect real applications of indirect questioning techniques as well. Even though comprehensibility in our study may actually have measured a mixture of comprehension and motivation, there is therefore little reason to expect a higher share of valid responses in real applications than in the present study. To further explore the exact mechanisms underlying incorrect responses, future studies should however try to measure task motivation more directly, or might try to increase task motivation by offering financial incentives.

As participants had to take on artificial characters' perspectives in a scenario-based design, absolute comprehension rates and perceived privacy scores might not be directly transferrable to real applications. However, if participants respond to sensitive questions from their own perspective, comprehension and perceived privacy protection are intertwined by

default. For example, carriers of a sensitive attribute who do not trust a questioning technique will necessarily tend to provide untruthful (that is, incorrect) responses; vice versa, carriers who fully trust the procedure will probably answer truthfully (that is, correctly). For this reason, only a scenario-based approach allows to separate comprehension from perceived privacy protection in RRT designs investigating sensitive attributes; and arguably, at least the rank order of the questioning techniques we investigated is therefore likely to remain valid even if absolute values may differ in real applications.

Another limitation of the present study is that we measured perceived privacy protection in a within-subjects design. While this may have affected responses, it allowed us to achieve higher statistical power, and also helped to avoid an effect that has been shown to potentially distort the results of between-subjects comparisons of numerical rating scales (Birnbbaum, 1999). In particular, contexts that differ between experimental conditions can lead to erroneous conclusions in between-subjects designs if participants provide relative judgments according to the range principle. For example, in a between-subjects design, participants have been shown to perceive the number 9 as being higher than the number 221 if the former evoked a frame of reference that consisted of single digit numbers, whereas the latter evoked a frame of reference that consisted of three-digit numbers (Birnbbaum, 1999). Similarly, an absolute judgment of the privacy protection afforded by a direct question may be distorted if participants are not aware of the possibility of privacy-protecting indirect questioning techniques because they are not given an opportunity to acquaint themselves with such techniques. Our decision to employ a within-subjects design helped to avoid such range effects because participants were given an opportunity to compare all questioning techniques.

A final limitation of our study is the relatively narrow age range of the participants (25 to 35 years). While this relatively homogeneous sample increased the statistical power to detect differences between experimental conditions, it also limits the generalizability of the findings. Future studies should therefore include older participants to investigate the replicability of our results in samples with a broader range of age.

This study supports the application of indirect questioning designs since they were shown to increase perceived privacy protection. When selecting among techniques, the best advice is to use CWM (Yu et al., 2008) to assess sensitive personal attributes. This model had the highest comprehensibility among indirect questioning techniques, and substantially increased perceived privacy protection in comparison to direct questioning. This recommendation is supported further by findings from various extant studies that suggest CWM results in more valid prevalence estimates than conventional direct questioning (e.g., Coutts et al., 2011; Hoffmann & Musch, 2015; Jann, Jerke, & Krumpal, 2012; Kundt et al., 2013; Nakhaee, Pakravan, & Nakhaee, 2013). If the attribute under investigation is extraordinarily sensitive (e.g., deviant sexual interests or severe criminal behavior), researchers may want to consider using the UCT (Miller, 1984) to maximize perceived privacy.

## References

- Abernathy, J. R., Greenberg, B. G., & Horvitz, D. G. (1970). Estimates of Induced Abortion in Urban North-Carolina. *Demography*, 7(1), 19-29.
- Abul-Ela, A.-L. A., Greenberg, B. G., & Horvitz, D. G. (1967). A Multi-Proportions Randomized Response Model. *Journal of the American Statistical Association*, 62, 990-1008.
- Ahart, A. M., & Sackett, P. R. (2004). A new method of examining relationships between individual difference measures and sensitive behavior criteria: Evaluating the unmatched count technique. *Organizational Research Methods*, 7, 101-114. doi: 10.1177/1094428103259557
- Baudson, T. G., & Preckel, F. (2015). mini-q: Intelligenzscreening in drei Minuten [mini-q: a three-minute intelligence screening]. *Diagnostica, Advance online publication*. doi: 10.1026/0012-1924/a000150
- Birnbaum, M. H. (1999). How to show that  $9 > 221$ : Collect judgments in a between-subjects design. *Psychological Methods*, 4, 243-249. doi: 10.1037/1082-989x.4.3.243
- Böckenholt, U., Barlas, S., & van der Heijden, P. G. M. (2009). Do Randomized-Response Designs Eliminate Response Biases? An Empirical Study of Non-Compliance Behavior. *Journal of Applied Econometrics*, 24, 377-392. doi: Doi 10.1002/Jae.1052
- Boeije, H. R., & Lensvelt-Mulders, G. J. L. M. (2002). Honest by chance: A qualitative interview study to clarify respondents (non-) compliance with computer-assisted randomized response. *Bulletin Methodologie Sociologique*, 75, 24-39.
- Chaudhuri, A., & Christofides, T. C. (2013). *Indirect Questioning in Sample Surveys*. Berlin, Heidelberg: Springer.

- Clark, S. J., & Desharnais, R. A. (1998). Honest answers to embarrassing questions: Detecting cheating in the randomized response model. *Psychological Methods, 3*, 160-168.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Coutts, E., & Jann, B. (2011). Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods & Research, 40*, 169-193. doi: 10.1177/0049124110390768
- Coutts, E., Jann, B., Krumpal, I., & Näher, A.-F. (2011). Plagiarism in Student Papers: Prevalence Estimates Using Special Techniques for Sensitive Questions. *Jahrbücher für Nationalökonomie Und Statistik, 231*, 749-760.
- Dawes, R. M., & Moore, M. (1980). Die Guttman-Skalierung orthodoxer und randomisierter Reaktionen [Guttman scaling of orthodox and randomized reactions]. In F. Petermann (Ed.), *Einstellungsmessung, Einstellungsforschung [Attitude measurement, attitude research]* (pp. 117–133). Göttingen: Hogrefe.
- Edgell, S. E., Duchan, K. L., & Himmelfarb, S. (1992). An Empirical-Test of the Unrelated Question Randomized-Response Technique. *Bulletin of the Psychonomic Society, 30*, 153-156.
- Edgell, S. E., Himmelfarb, S., & Duchan, K. L. (1982). Validity of Forced Responses in a Randomized-Response Model. *Sociological Methods & Research, 11*, 89-100. doi: 10.1177/0049124182011001005
- Erdfelder, E., & Musch, J. (2006). Experimental methods of psychological assessment. In M. Eid & E. Diener (Eds.), *Handbook of Multimethod Measurement in Psychology* (pp. 205-220). Washington, D.C.: American Psychological Association.

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.
- Fidler, D. S., & Kleinknecht, R. E. (1977). Randomized Response Versus Direct Questioning - 2 Data-Collection Methods for Sensitive Information. *Psychological Bulletin*, *84*, 1045-1049.
- Fox, J. A., & Tracy, P. E. (1980). The Randomized-Response Approach - Applicability to Criminal-Justice Research and Evaluation. *Evaluation Review*, *4*(5), 601-622. doi: 10.1177/0193841x8000400503
- Fox, J. A., & Tracy, P. E. (1986). *Randomized Response: A Method for Sensitive Surveys*. Beverly Hills, CA: Sage.
- Goodstadt, M. S., & Gruson, V. (1975). Randomized Response Technique - Test on Drug-Use. *Journal of the American Statistical Association*, *70*, 814-818.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, *59*, 93-104. doi: 10.1037/0003-066X.59.2.93
- Hejri, S. M., Zendehdel, K., Asghari, F., Fotouhi, A., & Rashidian, A. (2013). Academic disintegrity among medical students: a randomised response technique study. *Medical Education*, *47*, 144-153. doi: 10.1111/Medu.12085

- Hoffmann, A., Diedenhofen, B., Verschuere, B. J., & Musch, J. (2015). A strong validation of the Crosswise Model using experimentally induced cheating behavior. *Experimental Psychology*, *62*, 403-414. doi: 10.1027/1618-3169/a000304
- Hoffmann, A., & Musch, J. (2015). Assessing the validity of two indirect questioning techniques: a Stochastic Lie Detector versus the Crosswise Model. *Behavior Research Methods*, Advance online publication. doi: 10.3758/s13428-015-0628-6
- Holbrook, A. L., & Krosnick, J. A. (2010). Measuring Voter Turnout by Using the Randomized Response Technique: Evidence Calling into Question the Method's Validity. *Public Opinion Quarterly*, *74*, 328-343. doi: 10.1093/Poq/Nfq012
- Horvitz, D. G., Shah, B. V., & Simmons, W. R. (1967). The Unrelated Question Randomized Response Model. *Proceedings of the Social Statistics Section, American Statistical Association*.
- I-Cheng, C., Chow, L. P., & Rider, R. V. (1972). Randomized Response Technique as Used in Taiwan Outcome of Pregnancy Study. *Studies in Family Planning*, *3*, 265-269.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434-446. doi: 10.1016/j.jml.2007.11.007
- James, R. A., Nepusz, T., Naughton, D. P., & Petroczi, A. (2013). A potential inflating effect in estimation models: Cautionary evidence from comparing performance enhancing drug and herbal hormonal supplement use estimates. *Psychology of Sport and Exercise*, *14*, 84-96. doi: 10.1016/j.psychsport.2012.08.003
- Jann, B., Jerke, J., & Krumpal, I. (2012). Asking Sensitive Questions Using the Crosswise Model. *Public Opinion Quarterly*, *76*, 32-49. doi: 10.1093/Poq/Nfr036

- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47, 2025-2047. doi: 10.1007/s11135-011-9640-9
- Kulka, R. A., Weeks, M. F., & Folsom, R. E. (1981). *A comparison of the randomized response approach and direct questioning approach to asking sensitive survey questions*. Working paper. NC: Research Triangle Institute.
- Kundt, T. C., Misch, F., & Nerré, B. (2013). Re-assessing the merits of measuring tax evasions through surveys: Evidence from Serbian firms. ZEW Discussion Papers, No. 13-047. Retrieved Dec 12th, 2013, from <http://hdl.handle.net/10419/78625>
- Lamb, C. W., & Stem, D. E. (1978). An Empirical Validation of the Randomized Response Technique. *Journal of Marketing Research*, 15, 616-621. doi: Doi 10.2307/3150633
- Landsheer, J. A., van der Heijden, P. G. M., & van Gils, G. (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, 1-12. doi: 10.1023/A:1004361819974
- Lensvelt-Mulders, G. J. L. M., & Boeije, H. R. (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, 591-608. doi: 10.1016/j.chb.2004.11.001
- Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M., & Maas, C. J. M. (2005). Meta-analysis of randomized response research: thirty-five years of validation. *Sociological Methods & Research*, 33, 319-348. doi: 10.1177/0049124104268664
- Locander, W., Sudman, S., & Bradburn, N. (1976). An Investigation of Interview Method, Threat and Response Distortion. *Journal of the American Statistical Association*, 71, 269-275. doi: 10.2307/2285297

- Mangat, N. S. (1994). An Improved Randomized-Response Strategy. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *56*, 93-95.
- Mangat, N. S., & Singh, R. (1990). An Alternative Randomized-Response Procedure. *Biometrika*, *77*, 439-442. doi: 10.1093/biomet/77.2.439
- Marquis, K. H., Marquis, M. S., & Polich, J. M. (1986). Response Bias and Reliability in Sensitive Topic Surveys. *Journal of the American Statistical Association*, *81*, 381-389. doi: 10.2307/2289227
- Miller, J. D. (1984). *A new survey technique for studying deviant behavior*. Unpublished Ph.D. dissertation, George Washington University, Department of Sociology.
- Moshagen, M., Hilbig, B. E., Erdfelder, E., & Moritz, A. (2014). An Experimental Validation Method for Questioning Techniques That Assess Sensitive Issues. *Experimental Psychology*, *61*, 48-54. doi: 10.1027/1618-3169/a000226
- Moshagen, M., Musch, J., & Erdfelder, E. (2012). A stochastic lie detector. *Behavior Research Methods*, *44*, 222-231. doi: 10.3758/s13428-011-0144-2 21858604
- Moshagen, M., Musch, J., Ostapczuk, M., & Zhao, Z. (2010). Reducing Socially Desirable Responses in Epidemiologic Surveys. An Extension of the Randomized-response Technique. *Epidemiology*, *21*, 379-382. doi: 10.1097/Ede.0b013e3181d61dbc
- Musch, J., Bröder, A., & Klauer, K. C. (2001). Improving Survey Research on the World-Wide Web using the Randomized Response Technique. In U. D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 179-192). Lengerich, Germany: Pabst.
- Nakhaee, M. R., Pakravan, F., & Nakhaee, N. (2013). Prevalence of Use of Anabolic Steroids by Bodybuilders Using Three Methods in a City of Iran. *Addict Health*, *5*(3-4), 1-6.

- O'Brien, D. (1977). *The Comprehension Factor in Randomized Response*. Ph.D. thesis, University of Wyoming, Laramie, Wyoming.
- Ostapczuk, M., Moshagen, M., Zhao, Z., & Musch, J. (2009). Assessing sensitive attributes using the randomized response technique: Evidence for the importance of response symmetry. *Journal of Educational and Behavioral Statistics*, *34*, 267-287. doi: 10.3102/1076998609332747
- Ostapczuk, M., Musch, J., & Moshagen, M. (2009). A randomized-response investigation of the education effect in attitudes towards foreigners. *European Journal of Social Psychology*, *39*, 920-931. doi: 10.1002/ejsp.588
- Ostapczuk, M., Musch, J., & Moshagen, M. (2011). Improving self-report measures of medication non-adherence using a cheating detection extension of the randomised-response-technique. *Statistical Methods in Medical Research*, *20*, 489-503. doi: 10.1177/0962280210372843
- Scheers, N. J., & Dayton, C. M. (1987). Improved Estimation of Academic Cheating Behavior Using the Randomized-Response Technique. *Research in Higher Education*, *26*(1), 61-69. doi: 10.1007/Bf00991933
- Soeken, K. L., & Macready, G. B. (1982). Respondents Perceived Protection When Using Randomized-Response. *Psychological Bulletin*, *92*, 487-489.
- Tian, G.-L., & Tang, M.-L. (2014). *Incomplete Categorical Data Design: Non-Randomized Response Techniques for Sensitive Questions in Surveys*. Boca Raton, FL: CRC Press, Taylor & Francis Group.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*, 859-883. doi: 10.1037/0033-2909.133.5.859 17723033

- Ulrich, R., Schröter, H., Striegel, H., & Simon, P. (2012). Asking Sensitive Questions: A Statistical Power Analysis of Randomized Response Models. *Psychological Methods, 17*(4), 623-641. doi: Doi 10.1037/A0029314
- Umesh, U. N., & Peterson, R. A. (1991). A Critical Evaluation of the Randomized-Response Method - Applications, Validation, and Research Agenda. *Sociological Methods & Research, 20*, 104-138.
- van der Heijden, P. G. M., van Gils, G., Bouts, J., & Hox, J. J. (1998). A comparison of randomized response, CASAQ, and direct questioning; eliciting sensitive information in the context of social security fraud. *Kwantitatieve Methoden, 19*, 15-34.
- van der Heijden, P. G. M., van Gils, G., Bouts, J., & Hox, J. J. (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 & Research, 28*, 505-537.
- Warner, S. L. (1965). Randomized-Response - a Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association, 60*, 63-69.
- Wimbush, J. C., & Dalton, D. R. (1997). Base rate for employee theft: Convergence of multiple methods. *Journal of Applied Psychology, 82*, 756-763.
- Wolter, F., & Preisendörfer, P. (2013). Asking Sensitive Questions: An Evaluation of the Randomized Response Technique Versus Direct Questioning Using Individual Validation Data. *Sociological Methods & Research, 42*, 321-353. doi: 10.1177/0049124113500474
- Yu, J.-W., Tian, G.-L., & Tang, M.-L. (2008). Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika, 67*, 251-263. doi: 10.1007/s00184-007-0131-x

## Tables

Table 1

*Nonparametric correlation coefficients (Spearman's Rho) measuring the association of comprehension and perceived privacy protection.*

Group	Questioning Technique				
	DQ (control)	CDM	SLD	CWM	UCT
Total sample ( $N = 401$ )	-.08	-.06	.04	.02	.09
High education ( $N = 202$ )	-.12	.04	.01	-.003	.12
Low education ( $N = 199$ )	-.02	-.12	.09	.07	.04

*Note.* DQ = Direct Question, CDM = Cheating Detection Model, SLD = Stochastic Lie Detector, CWM = Crosswise Model, UCT = Unmatched Count Technique. No correlation was statistically significant (all  $p > .05$ ).

## Figures

In the following, you will be required to respond to a question regarding academic dishonesty. If you were born in November or December, please answer “yes”, regardless of your true answer. If you were born in any other month, please answer truthfully.

**Question:** Have you ever cheated on an exam?

Yes       No

*Figure 1.* Example of a question regarding academic dishonesty as presented in surveys employing the Cheating Detection Model (Clark & Desharnais, 1998). The respondent’s month of birth is used as a randomization device with randomization probability  $p = 2/12 = .17$ .

In the following, you will be presented with two oppositional questions regarding academic dishonesty. If you have ever cheated on an exam before, please respond to question A. If you have never cheated on an exam before, please respond to...

- question A if you were born in November or December,
- question B if you were born in any other month.

**Question A:** Have you ever cheated on an exam?  
**Question B:** Have you never cheated on an exam?

Yes       No

*Figure 2.* Example of a question regarding academic dishonesty using the Stochastic Lie Detector (Moshagen et al., 2012). The respondent's month of birth is used as a randomization device with randomization probability  $p = 2/12 = .17$ .

In the following, you will be presented with two questions simultaneously, one regarding academic dishonesty, and the other regarding your month of birth.

**Question A:** Have you ever cheated on an exam?

**Question B:** Were you born in November or December?

Yes to both questions or no to both questions

Yes to exactly one of the questions (regardless of which one)

*Figure 3.* Example of a question regarding academic dishonesty using the Crosswise Model (Yu et al., 2008). The respondent's month of birth is used as a randomization device with randomization probability  $p = 2/12 = .17$ .

In the following, you will be presented with four questions simultaneously. Please indicate your total number of “Yes”-responses, regardless of your individual answers.

**Question A:** Have you ever cheated on an exam?  
**Question B:** Were you born in November or December?  
**Question C:** Are you a male?  
**Question D:** Have you ever been to the city of London?

Total number of “Yes”-responses (0 to 4): \_\_\_\_\_

*Figure 4.* Example of a question regarding academic dishonesty using the Unmatched Count Technique (Miller, 1984) with one sensitive (A) and three non-sensitive questions (B to D).

In the following, you will be required to respond to a question from the perspective of Wilhelm.



**Wilhelm**

- Wilhelm has never cheated on an exam,
- Wilhelm was born in July,
- Wilhelm is a man,
- Wilhelm has never been to London.

Please read the instructions carefully and try to find out which answer Wilhelm would have to give:

In the following, you will be presented with two statements A and B. You will not be required to respond to each statement separately, but only to indicate whether...

- Both statements or none of the two statements are true,

**or**

- Exactly one statement (irrespective of which one) is true.

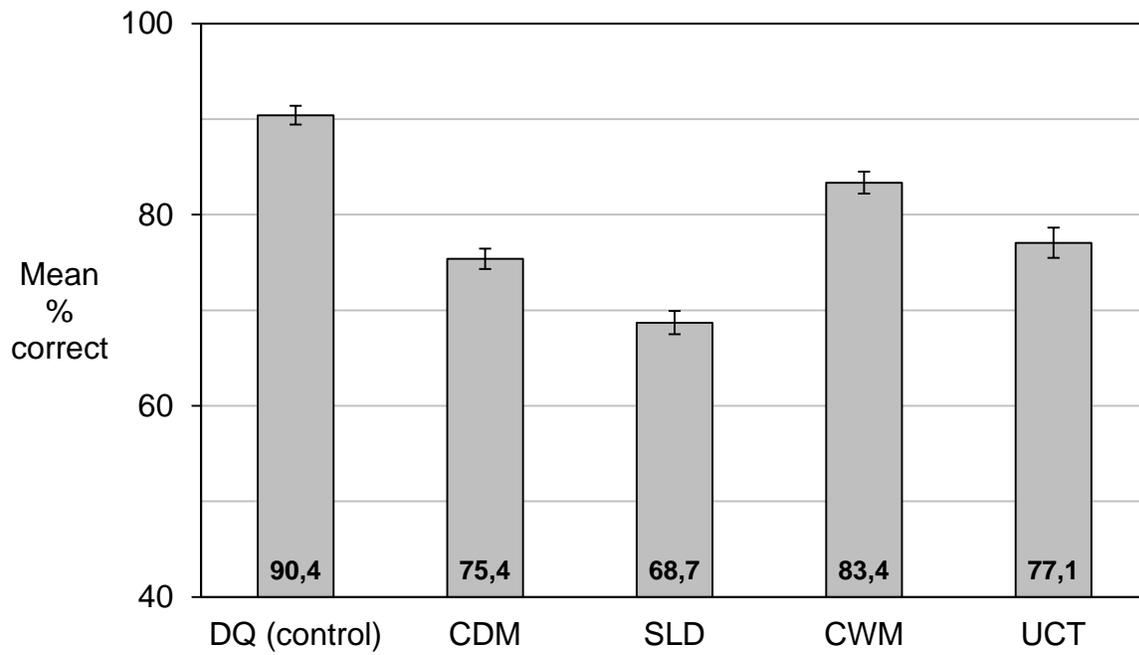
As we do not know your single answers to each of the statements, your response will remain confidential.

**Statement A:** "I have cheated on an exam."  
**Statement B:** "I was born in November or December."

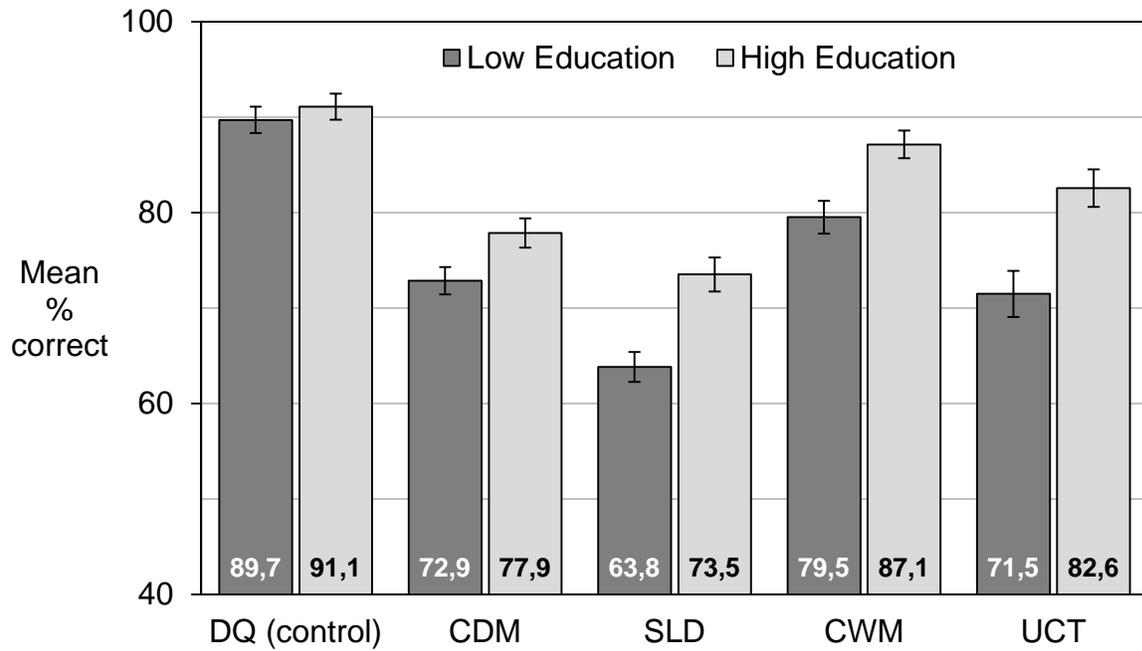
Which answer would Wilhelm have to give?

- "Yes to both questions or no to both questions."
- "Yes to exactly one of the questions (regardless of which one)."

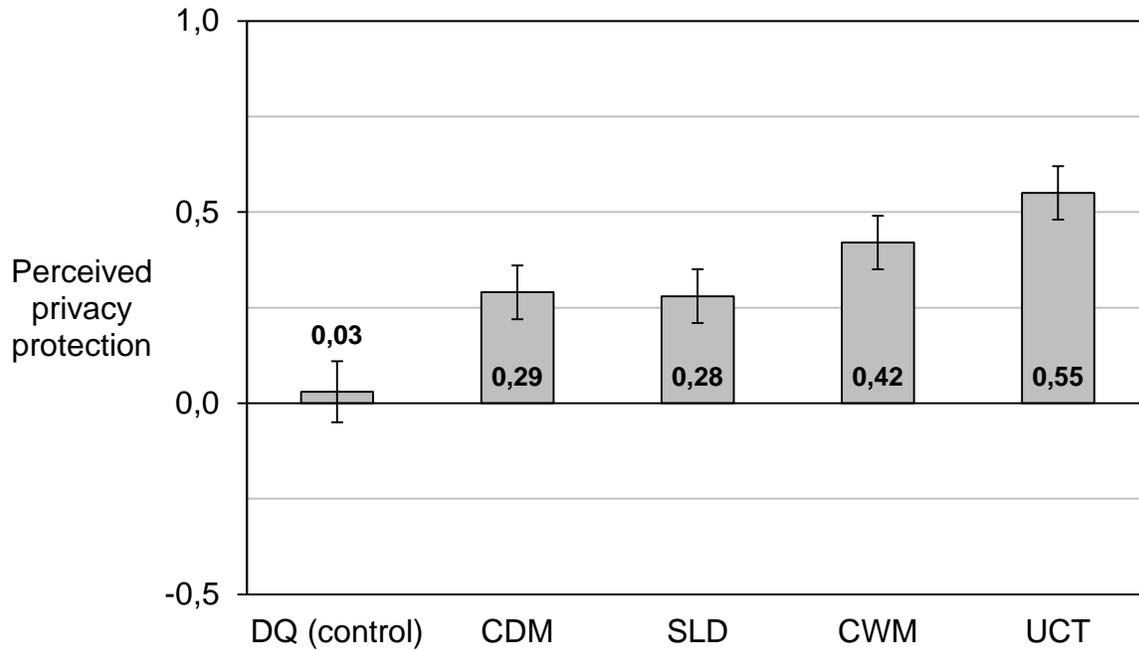
*Figure 5.* Screenshot of a CWM question that had to be answered from the perspective of the fictional character Wilhelm. As Wilhelm never cheated on an exam and was born in July, the first answer option ("Yes to both questions or no to both questions.") would have been correct.



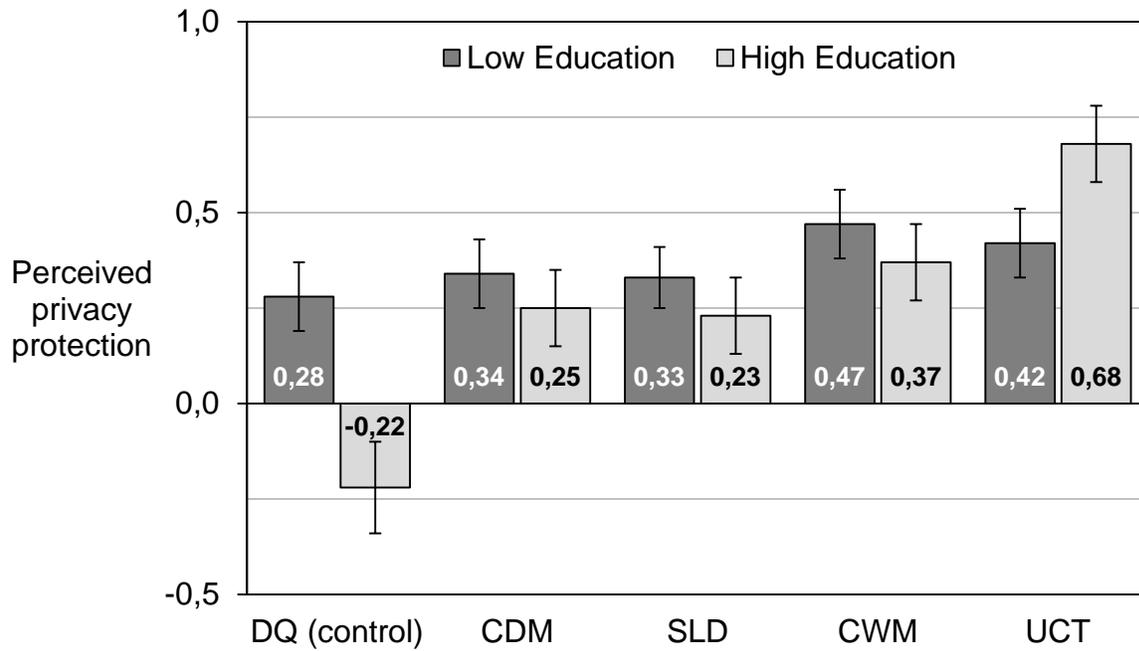
*Figure 6.* Mean percent of correct responses as a function of questioning technique in the total sample ( $N = 401$ ). Error bars denote  $\pm 1$  standard error.



*Figure 7.* Mean percent of correct responses as a function of questioning technique and low ( $N = 199$ ) versus high education ( $N = 202$ ). Error bars denote  $\pm 1$  standard error.



*Figure 8.* Mean perceived privacy protection on a 7-point Likert-scale from -3 (no confidentiality) to +3 (perfect confidentiality) as a function of questioning technique in the total sample ( $N = 401$ ). Error bars denote +/- 1 standard error.



*Figure 9.* Mean perceived privacy protection on a 7-point Likert-scale from -3 (no confidentiality) to +3 (perfect confidentiality) as a function of questioning technique and low ( $N = 199$ ) versus high education ( $N = 202$ ). Error bars denote +/- 1 standard error.