A strong validation of the Crosswise Model using experimentally induced cheating behavior

Hoffman, A.; Diedenhofen, B.; Verschuere, B.; Musch, J.

DOI
10.1027/1618-3169/a000304

Publication date
2015

Document Version
Final published version

Published in
Experimental Psychology

Citation for published version (APA):
Hoffman, A., Diedenhofen, B., Verschuere, B., & Musch, J. (2015). A strong validation of the Crosswise Model using experimentally induced cheating behavior. *Experimental Psychology*, *62*(6), 403-414. https://doi.org/10.1027/1618-3169/a000304

Volume 62 / Number 6 / 2015
ISSN-L 1618-3169 · ISSN-Print 1618-3169 · ISSN-Online 2190-5142

Experimental Psychology

6/15

www.hogrefe.com/journals/exppsy

HOGREFE

Contents

Experimental Psychology

HOGREFE

Short Research Article

A Strong Validation of the Crosswise Model Using Experimentally-Induced Cheating Behavior

Adrian Hoffmann,[1] Birk Diedenhofen,[1] Bruno Verschuere,[2] and Jochen Musch[1]

[1]Department of Experimental Psychology, University of Düsseldorf, Germany, [2]Department of Clinical Psychology, University of Amsterdam and Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands; Department of Psychology, Ghent University, Belgium

Abstract. We constructed an online cheating paradigm that could be used to validate the Crosswise Model (Yu, Tian, & Tang, 2008), a promising indirect questioning technique designed to control for socially desirable responding on sensitive questions. Participants qualified for a reward only if they could identify the target words from three anagrams, one of which was virtually unsolvable as shown on a pretest. Of the 664 participants, 15.5% overreported their performance and were categorized as cheaters. When participants were asked to report whether they had cheated, a conventional direct question resulted in a substantial underestimate (5.1%) of the known prevalence of cheaters. Using a CWM question resulted in a more accurate estimate (13.0%). This result shows that the CWM can be used to control for socially desirable responding and provides estimates that are much closer to the known prevalence of a sensitive personal attribute than those obtained using a direct question.

Keywords: Crosswise Model, randomized response technique, strong validation, cheating

On surveys that ask for direct self-reports to assess sensitive personal attributes, some individuals tend to refrain from answering truthfully in order to present themselves in a socially desirable light (Paulhus, 1991; Tourangeau & Yan, 2007). Such socially desirable responding may result in biased prevalence estimates of socially desirable or undesirable attitudes and behaviors (Krumpal, 2013; Phillips & Clancy, 1972). The validity of prevalence estimates is particularly threatened when employing a question that directly asks whether the respondent embodies the sensitive attribute (*direct questioning*; DQ).

Scientific research has spawned various approaches that aim to measure or control for the influence of the social desirability bias (Nederhof, 1985; Paulhus, 1991). To increase the validity of prevalence estimates based on self-reports, indirect questioning procedures such as the Randomized Response Technique (RRT; Warner, 1965) have been proposed. In the original RRT procedure, respondents are instructed to answer either a positively ("Do you embody the sensitive attribute?") or negatively worded sensitive question ("Do you not embody the sensitive attribute?"), depending on the outcome of an external randomization process (e.g., the cast of a die). The individual outcome of the randomization process remains unknown to the questioner, but the distribution of the randomization outcomes and thus the probability of selecting either of the questions are known at sample level. Therefore, neither a "yes" nor a "no" response will allow any conclusions to be drawn about an individual respondent's true status with respect to the sensitive attribute. The perceived confidentiality of responses and the respondent's probability of answering truthfully is thereby increased. Despite the increased confidentiality at the level of the individual, the known distribution of the randomization outcome allows an estimate to be made of the prevalence of a sensitive attribute in the sample with presumably less of a bias from socially desirable responding. However, unsystematic variance is added to individual responses; estimates obtained using the RRT therefore suffer from increased variance compared with more conventional direct questions. This decrease in efficiency is supposed to be compensated for by an increase in validity, especially if the attribute in question is highly sensitive in nature (Lensvelt-Mulders, Hox, van der Heijden, & Maas, 2005). However, a loss in efficiency is considered tolerable only to the extent to which the use of an indirect question actually results in more valid prevalence estimates than the use of a direct question.

As a promising new approach for controlling for socially desirable responding, Yu, Tian, and Tang (2008)

DOI: 10.1027/1618-3169/a000304

proposed the Crosswise Model (CWM). This technique allows questions to be asked indirectly without requiring participants to operate an external randomization device (e.g., a die). Questions asked in the CWM format require participants to respond to two statements simultaneously. One of these statements asks for the sensitive attribute ("I embody the sensitive attribute X") with unknown prevalence π; the other statement asks about a second nonsensitive attribute for which the prevalence p is known in the population (e.g., "I was born in November or December"). Participants are then instructed to indicate whether (a) "either statement or neither of the two statements is true," or whether (b) "exactly one of the two statements (regardless of which one) is true." Neither of the two answer options (a) or (b) allows inferences to be made about a respondent's individual status with regard to the sensitive attribute. For the entire sample, however, Yu et al. (2008) showed that π can be estimated by

$$\hat{\pi} = \frac{p - 1 + \frac{n'}{n}}{2p - 1}, p \neq 1/2 \quad (1)$$

with n' representing the total number of "both true/both false" responses and n reflecting the sample size. An estimate of the variance of $\hat{\pi}$ is given by

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

Whereas the CWM is mathematically equivalent to the original RRT model proposed by Warner (Ulrich, Schröter, Striegel, & Simon, 2012), its questioning format offers potential advantages over previous approaches. As the randomization procedure is integrated into the question itself, the instructions are simpler than in competing RRT procedures; this makes the CWM presumably "easy to operate for both interviewer and interviewee" (Yu et al., 2008, p. 255). Therefore, Tian and Tang (2014) suggested that the CWM be referred to as a "Nonrandomized Response Technique." Indeed, a recent study found the CWM to be superior to various competing indirect questioning techniques with regard to comprehensibility and perceived privacy protection (Hoffmann, Schmidt, Waubert de Puiseau, & Musch, 2015). Moreover, the CWM offers response symmetry in the sense that none of the answer options offers a "safe alternative" to which respondents might turn to dispel any connection to the sensitive attribute. Response symmetry has been shown to increase compliance with instructions and hence to increase the validity of prevalence estimates (Ostapczuk, Moshagen, Zhao, & Musch, 2009). However, whether the use of CWM questions does, indeed, lead to more valid prevalence estimates has yet to be shown in empirical validation studies.

Validation studies for indirect questioning techniques can be roughly divided into two categories (Moshagen, Hilbig, Erdfelder, & Moritz, 2014): "Weak" validation studies compare prevalence estimates obtained with different questioning techniques (e.g., direct vs. indirect questions). For socially undesirable attributes, higher prevalence estimates are usually considered to have higher validity, assuming that the social desirability bias leads to underestimates of the true value ("more-is-better" criterion; Umesh & Peterson, 1991). However, higher prevalence estimates may still under- or overestimate the true rate at which people embody a sensitive attribute if the prevalence in the sample and the prevalence in the population ("ground truth") remain unknown. Consequently, such studies can offer only weak evidence for the validity of a questioning technique. By contrast, "strong" validation studies rely on a sensitive attribute for which the true prevalence in the sample is known. If a questioning technique provides a prevalence estimate close to the known prevalence in the sample, this accordance is considered particularly strong evidence for its validity. Strong validation studies are considered the gold standard in the evaluation of methods that aim to control for social desirability bias (Lensvelt-Mulders et al., 2005; Moshagen et al., 2014). Unfortunately, the assessment of the true proportion of respondents in a sample who carry a sensitive attribute is usually costly and often impossible. For example, assessing a sample of individuals known to have been convicted of social welfare fraud (van der Heijden, van Gils, Bouts, & Hox, 2000) is possible only if legal regulations grant researchers access to the respective contact data. Moreover, accessing such sensitive information consumes time and money and may raise some ethical concerns because the individual status of respondents with respect to the sensitive attribute will be uncovered. Consequently, strong validation studies implementing indirect questioning techniques are very scarce; in Lensvelt-Mulders et al.'s (2005) meta-analysis, only six of 38 validation studies met this criterion. With respect to the CWM, some weak validation studies using the "more-is-better" approach have shown that the application of the CWM results in higher prevalence estimates than a conventional direct question for plagiarism in student papers (Jann, Jerke, & Krumpal, 2012), tax evasion (Korndörfer, Krumpal, & Schmukle, 2014; Kundt, Misch, & Nerré, 2013), steroid use (Nakhaee, Pakravan, & Nakhaee, 2013), and xenophobia and islamophobia (Hoffmann & Musch, 2015). However, only a strong validation study using a sensitive attribute with known prevalence can provide firm evidence for the validity of prevalence estimates obtained using the CWM. To the best of our knowledge, a strong validation study has yet to be conducted to investigate the CWM.[1]

As was recently shown by Moshagen et al. (2014), a promising approach that can be applied to overcome the notorious difficulties associated with strong validation studies is the experimental induction of a socially undesirable attribute. Moshagen et al.'s participants were instructed to secretly roll a die and to report the outcome, under the condition that certain outcomes were financially rewarded (an adjusted "die-under-the-cup paradigm"; e.g., Hilbig & Hessler, 2013). The proportion of respondents claiming

[1] An experiment pursuing a related approach was recently presented by Hoeglinger, Diekmann, and Jann (2014).

a reward was about twice as high as would be expected if all participants had honestly reported the outcome of their die roll. Consequently, about half of the respondents claiming a reward had to be categorized as cheaters (Moshagen et al., 2014). Subsequently, respondents were presented with a question about their cheating behavior either in direct questioning format, or in the format of the "Stochastic Lie Detector" (SLD; Moshagen, Musch, & Erdfelder, 2012). In this variant of the RRT, an additional parameter t is introduced to account for the proportion of carriers of the sensitive attribute responding truthfully, while the remaining carriers $(1 - t)$ are assumed to be lying to conceal their true status. While a direct question substantially underestimated the actual rate of cheaters, the application of the SLD resulted in a prevalence estimate that was fairly close to the true prevalence. Therefore, the SLD was evaluated favorably as a method that was capable of controlling for socially desirable responding (Moshagen et al., 2014).

Whereas the rationale of the die-under-the-cup task appears compelling, the procedure requires space, material, preparation, and a personal interaction with every respondent, making it difficult to employ in an online setting (but see, e.g., Shalvi, Dana, Handgraaf, & De Dreu, 2011). Online surveys, however, offer an attractive environment for indirect questioning studies as the large sample sizes that are required to compensate for the increased variance in indirect prevalence estimates can be efficiently assessed using the World Wide Web (Musch, Bröder, & Klauer, 2001). Therefore, we used a validation method similar to the die-under-the-cup paradigm presented in Moshagen et al. (2014) but developed it further into a format that could be employed in online assessments. Our approach was based on the "word-jumble task" proposed by Wiltermuth (2011). In this paradigm, participants are presented with a list of nine words with a jumbled letter sequence. They are instructed to rearrange the letter sequence of each jumble in their minds until the letters form a valid word; instead of having to report the solution, however, they are asked merely to report the number of consecutive jumbles they were able to solve from the beginning. They are told that a higher reported number of successively solved jumbles will be remunerated with a higher financial reward. However, the third consecutive anagram is chosen to be practically unsolvable. On pretests of several studies using the word-jumble task, 0 out of 10 (Halevy, Shalvi, & Verschuere, 2013), 0 out of 30 (Wiltermuth, 2011), and 0 out of 42 (Gino & Mogilner, 2014) participants were able to identify the solution of the third anagram. Hence, any participant reporting to have solved three or more consecutive scrambled words is most likely a cheater. In the original study by Wiltermuth (2011), about 29% of the respondents claimed to have solved three or more anagrams in a row. Remarkably, participants "perceived the act of over-reporting [their] performance as unethical and greedy" (p. 162), strongly indicating the socially undesirable nature of such behavior. Other studies reported cheating rates of 23% (Halevy et al., 2013), 52% (Ruedy, Moore, Gino, & Schweitzer, 2013), and 40%–73% under various experimental conditions (Gino & Mogilner, 2014). These results suggest that the word-jumble task is reliably capable of experimentally generating a substantial rate of cheaters in a given sample, even though cheating on the task is perceived as socially undesirable by the participants.

In the present study, we pursued two goals: First, on the basis of a variant of the anagram-cheating task proposed by Wiltermuth (2011), we constructed an experimental procedure that allowed for an online application of the validation method used by Moshagen et al. (2014). Second, we conducted the first strong validation of the CWM (Yu et al., 2008) using cheating behavior induced via the anagram-cheating task as an external criterion. Because dark personality traits such as psychopathy have been discussed as potential moderators of cheating behavior (Halevy et al., 2013), we included the "dirty dozen" items (Jonason & Webster, 2010; Küfner, Dufner, & Back, 2014) as brief measures of psychopathy, Machiavellianism, and narcissism.

Method

Participants and Design

A total of 698 registered members of a noncommercial research panel run by scientific employees of the University of Düsseldorf were recruited via email invitation and accessed the online questionnaire. Pretest participants were not eligible for the main study. Respondents who did not complete the experiment (34) were excluded from the analyses. This dropout was nonselective by experimental condition (10 of 138 [6.8%] participants within DQ condition; 24 of 526 [4.4%] participants within CWM condition; $\chi^2(1) = 1.44$, $p = .23$). The final sample of $N = 664$ participants (95.1% of the participants who had accessed the survey) had a mean age of 36.7 years ($SD = 13.73$, Min = 18, Max = 94); 380 (57.2%) of the participants were female.

Using a single-factor between-subjects design, we randomly assigned participants to respond to a question about their cheating behavior either in a DQ ($n = 138$; 20.8% of the final sample) or CWM format ($n = 526$). A skewed allocation probability of 1:4 was chosen to compensate for the comparably low efficiency of prevalence estimates obtained via the CWM (Ulrich et al., 2012).

Measures

Anagram-Cheating Task

Previous versions of the word-jumble task had been constructed and tested only in English (Gino & Mogilner, 2014; Ruedy et al., 2013; Wiltermuth, 2011) or Dutch (Halevy et al., 2013). We therefore created a new set of anagrams from scratch and conducted a first pretest to identify German items with suitable item difficulties. Our objective was to find two anagrams that would be solved by virtually every participant and one anagram that was practically

Table 1. Item characteristics of the three anagrams that were chosen for the final anagram-cheating task obtained in the first pilot study (N = 136)

	Target word	Anagram	Translation	N solved (%)	M_{pt} (SD)	M_{rt} (SD)
1	PRAXIS	APRXIS	Practice	136 (100.0)	1.99 (1.36)	3.51 (1.43)
2	UMWELT	UMELTW	Environment	136 (100.0)	1.90 (0.93)	3.09 (1.36)
3	TRIOLE	IERTLO	Triplet	0 (0.0)	18.73 (3.60)	5.30 (5.90)

Note. M_{pt} = mean presentation time in s; M_{rt} = mean response time in s.

Table 2. Item characteristics of the three anagrams that were chosen for the final anagram-cheating task obtained in the second pilot study (N = 183)

	Target word	Anagram	Translation	N solved (%)	M_{pt} (SD)	M_{rt} (SD)
1	PRAXIS	APRXIS	Practice	182 (99.5)	2.58 (1.66)	4.56 (1.69)
2	UMWELT	UMELTW	Environment	181 (98.9)	2.51 (2.11)	4.18 (1.90)
3	TRIOLE	IERTLO	Triplet	2 (1.1)	17.39 (5.03)	9.63 (7.06)

Note. M_{pt} = mean presentation time in s; M_{rt} = mean response time in s.

unsolvable. To this end, we constructed 32 presumably very easy anagrams, and 13 presumably very difficult anagrams. Details on the construction process along with a full list of all 45 anagrams employed in the first pilot study are presented in Appendix A. These anagrams were pretested on 136 individuals recruited online via the noncommercial research panel run by members of the Department of Experimental Psychology at the University of Düsseldorf. Items were presented in random order. For each anagram, participants were given a maximum of 20 s to identify the target word. If participants identified the target word, they were asked to press a button labeled "I found the target word and want to enter the solution." After the 20 s presentation time had elapsed or when the respective button was pressed, the anagram was masked and respondents were given another 20 s to enter the solution into an input box. Answers were forced to have exactly six letters and were scored as correct only when they provided a perfect match with the target word. Three anagrams had item characteristics that perfectly met the requirements of the cheating task (see Table 1): Two very easy anagrams were solved by every single participant, and one difficult anagram was solved by none of the 136 respondents.

These three anagrams were included in a second pretest on 183 participants recruited via the same online panel used in the first pretest. Instructions were widely identical to the first pilot study: Participants were given a maximum of 20 s per anagram to identify the target word, and another 20 s to enter the solution into an input box on a subsequent page. However, in line with the envisaged instructions in the main study, the anagrams were presented in a fixed order: The two very easy anagrams "APRXIS" and "UMELTW" were presented first, followed by the virtually unsolvable anagram "IERTLO." Additionally, following the paradigm proposed by Moshagen et al. (2014), we offered an incentive for solving all three anagrams (a chance to partake in a lottery for three gift certificates worth 20€, 30€, and 100€), expecting that this would increase participants' motivation to reach the maximum possible score. The second pretest confirmed the results of the first pretest (see Table 2): the two easy items were solved by almost all participants and the difficult item was virtually unsolvable.

In the final anagram-cheating task and as in the original paradigm used by Wiltermuth (2011), we simply asked participants to indicate the number of anagrams they were able to unscramble. Deviating from the original paradigm, and in line with our second pretest, however, only three anagrams were presented in a fixed order, and respondents were not asked to consider whether they had solved the anagrams successively. The three anagrams were presented for a maximum of 20 s each; participants could abort the presentation of each anagram if they believed they had identified the solution. On the basis of the pretest results, we expected that all or virtually all participants (approx. 99%) would be able to unscramble two of the anagrams but that virtually no participant (approx. 1%) would be able to identify all three target words. With a very low expected rate of false alarms, respondents claiming to have solved all three anagrams were therefore categorized as cheaters. In the style of our second pretest, we offered an incentive for claiming that all three anagrams were solved (see below), expecting that this would motivate a substantial number of participants to cheat knowingly.

Direct Question

Participants in the DQ condition were directly questioned about whether the following statement was true or false: "On the anagram task, I claimed that I had solved more anagrams than I had actually solved."

Crosswise Model

If assigned to the CWM condition, participants were simultaneously confronted with two statements. The wording of the sensitive statement was identical to the DQ condition. A second nonsensitive statement read: "I was born in November or December." The probability of being born in these months is about 15.8% according to official birth statistics provided by the German Federal Agency for Statistics (cf. Moshagen et al., 2012). Respondents were instructed to choose one of the two available answer options that stated (a) "Either statement or neither of the two statements is true" versus (b) "Exactly one statement (regardless of which one) is true," respectively. Additionally, participants were provided with a brief explanation of the rationale of the CWM. We ensured the respondents that their answers would not allow us to draw any conclusions regarding their true status with respect to the sensitive attribute, and would not have any effect on their chances to win in the lottery (see Appendix B for a complete reproduction of the CWM question).

Procedure

The study was administered as an online questionnaire using the EFS survey 10.2 software (Questback, 2014). The first page welcomed participants and obtained their informed consent. After participants were asked to provide demographic information, they were given the instructions for the anagram-cheating task and were informed that they would be given the chance to partake in a lottery for three gift certificates worth 20€, 30€, and 100€ if they performed well. Before the anagram-cheating task began, respondents completed two quite easy example anagrams to familiarize themselves with the paradigm. After completing the anagram-cheating task, respondents reported on a separate page how many anagrams they had solved. In the upper part of this page, all three anagrams were displayed again, along with their respective target words. In the middle part, participants were reminded that only a high performance would allow them to participate in the lottery. At the lower end of the page, the question about their performance on the anagram task asked: "How many of the anagrams did you solve in the available time?" The answer options were: "I did not solve any of the three anagrams," "I solved one of the three anagrams," "I solved two of the three anagrams," and "I solved all three anagrams." Only the last of these potential answers was identified as allowing the participant to partake in the lottery at the end of the survey. On the next page, depending on the experimental condition, a question in either a DQ or CWM format asked about whether respondents had just cheated on the anagram task by overreporting their results. Subsequently, the "dirty dozen" items (Jonason & Webster, 2010; Küfner et al., 2014) were presented on a single page to assess the dark personality constructs of psychopathy, Machiavellianism, and narcissism. On the last two pages, participants were debriefed, thanked for their cooperation, and offered the opportunity to enter the lottery. Because no one should be discriminated against for being honest, we offered all participants the opportunity to participate in the lottery regardless of their performance.

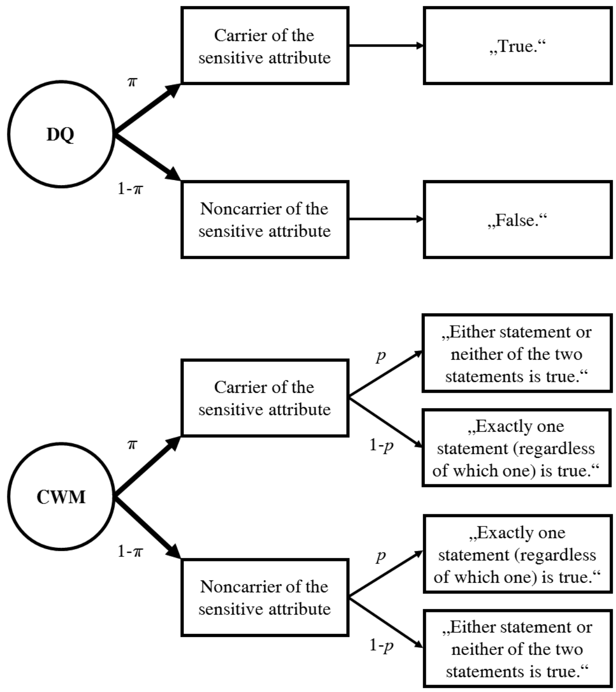

Figure 1. Combined multinomial processing tree model of the DQ and CWM questioning conditions.

Statistical Analysis

In the CWM condition, an estimate for the prevalence of cheating can be obtained by using Equations 1 and 2. In the DQ condition, the proportion of respondents answering "true" to a direct question provides a direct estimate of the prevalence of cheating. Following previous studies implementing indirect questioning techniques (e.g., Moshagen et al., 2012; Ostapczuk et al., 2009; Ostapczuk & Musch, 2011), however, we translated both the DQ and CWM conditions into a combined multinomial processing tree (MPT) model (Batchelder, 1998; Hu & Batchelder, 1994; see Figure 1).

This approach offers more flexibility in parameter estimation and offers convenient statistical tests of parameter restrictions (Moshagen et al., 2012). We obtained estimates for the prevalence of cheating in the DQ and CWM conditions via maximum likelihood procedures using the *multiTree* v0.41 software (Moshagen, 2010). To compare parameter estimates with each other and with the known true proportion of cheaters, differences in model fit between an unrestricted baseline model and a restricted alternative model (e.g., in which the DQ and CWM prevalence estimates were equalized) were assessed via differences in the asymptotically χ^2-distributed log-likelihood ratio statistic G^2 (Read & Cressie, 1988).[2]

Results

Cheating Behavior

Of the 664 participants, 103 (15.5%) claimed that they had solved all three anagrams. As the probability of achieving a score of three anagrams had been shown to be close to zero on both pretests, we used this percentage as a proxy for the "true" prevalence of cheating in our sample. Of the remaining 561 respondents who were classified as noncheaters, 558 (99.5% of the noncheaters, 84.0% of the sample) reported that they had solved exactly two anagrams, whereas only two participants (0.4% of the noncheaters, 0.3% of the sample) indicated that they had solved one anagram, and one participant (0.18% of the noncheaters, 0.15% of the sample) reported being unable to unscramble any words. These findings confirmed the results of our pretests and met our expectations that virtually all participants would solve two anagrams but that some would choose to overreport their performance to maximize their personal benefit.

Prevalence Estimates

Empirically observed answering frequencies by experimental condition that were used for parameter estimation are given in Table 3. All data sets and script files neccessary to reproduce the analyses reported for both pretests and the main study are provided as Electronic Supplementary Materials, ESM1–9 (see below).

As Table 4 shows, the application of a direct question resulted in a significant underestimation (5.1%) of the known "true" prevalence of cheating (15.5%) by about 10.4 percentage points, ΔG^2 ($df = 1$) = 14.88, $p < .001$. Thus, a substantial share of the cheaters in the DQ condition apparently refrained from telling the truth, perhaps to conceal that they had engaged in a socially undesirable behavior. By contrast, the prevalence estimate obtained via the indirect question in the CWM format (13.0%) deviated by only about 2.5 percentage points from the "true" prevalence, and this small difference was not significantly different from zero, ΔG^2 ($df = 1$) = 0.79, $p = .37$. Thus, the CWM seemed to be able to obtain an unbiased estimate for the prevalence of cheating in the sample. Finally, the estimates obtained in the DQ and CWM conditions (5.1% vs. 13.0%) differed significantly from each other, ΔG^2 ($df = 1$) = 5.29, $p < .05$.[3]

Table 3. Observed answering frequencies by experimental condition used for parameter estimation

	n (%)
DQ (*n* = 138)	
"True."	7 (5.1)
"False."	131 (94.9)
CWM (*n* = 526)	
"Either statement or neither of the two statements is true."	396 (75.3)
"Exactly one statement (regardless of which one) is true."	130 (24.7)

Note. DQ = direct question; CWM = Crosswise Model.

Discussion

The present study aimed to achieve two objectives: First, we wanted to develop an online paradigm that could induce a socially undesirable attribute and provide information about whether a participant embodied this attribute. To this end, we created an anagram task that incited participants to overreport their performance, thus inducing undesirable cheating behavior. In a second step, we used this anagram-cheating task to conduct a strong validation of the Crosswise Model (CWM; Yu et al., 2008), an indirect questioning technique designed to control for socially desirable responding. The known rate of cheaters in the sample served as an objective external criterion for the validity of the CWM prevalence estimate. In light of the present results, the application of the anagram-cheating task seemed to have successfully motivated a substantial proportion of participants (about 15.5%) to engage in a socially undesirable behavior. When responding to a conventional direct question, only 5.1% of the respondents honestly admitted their cheating. We attribute this difference between the levels of reported and actual behavior to the influence of social desirability bias. Hence, our results

[2] The MPT approach is only one of several methods to estimate the prevalence of sensitive attributes based on empirically observed answering frequencies in indirect questioning conditions. Alternatively, prevalence estimates can be obtained using simple closed-form calculations, and estimates can be compared to each other or to known values by conducting standard z-tests (as, e.g., demonstrated in Moshagen et al., 2014).

[3] None of the three dark personality traits we assessed (psychopathy, Machiavellianism, and narcissism) showed any association with cheating behavior or respondents' answering behavior in the DQ condition. Therefore, we will not report them in detail. However, for the purpose of future meta-analyses and to avoid a potential publication bias, we will report the correlation coefficients we found: psychopathy and cheating ($N = 644$): $r = .04$, $p = .33$; Machiavellianism and cheating ($N = 644$): $r = .07$, $p = .10$; narcissism and cheating ($N = 644$): $r = .03$, $p = .52$. In the DQ condition, we also found no significant correlations between psychopathy and lying ($N = 137$): $r = .09$, $p = .32$; Machiavellianism and lying ($N = 137$): $r = .07$, $p = .42$, and narcissism and lying ($N = 137$): $r < .01$, $p > .99$.

Table 4. Parameter estimates (standard errors in parentheses) for the prevalence of cheating on the anagram task

		Test against "true" prevalence of 15.5%		Test against CWM prevalence estimate	
Mode	$\hat{\pi}$	ΔG^2 (df = 1)	p	ΔG^2 (df = 1)	p
DQ (n = 138)	5.1% (1.9)	14.88	< .001	5.29	< .05
CWM (n = 526)	13.0% (2.8)	0.79	.37	–	

Note. $\hat{\pi}$ = Prevalence of cheating as estimated using a direct question and the Crosswise Model.

further support the notion that the assessment of sensitive attributes via direct self-reports may lead to an underestimation of their prevalence, as some carriers of the sensitive characteristic tend to conceal their true status (Krumpal, 2013; Phillips & Clancy, 1972). In the CWM condition, however, a fairly accurate estimate of the prevalence of cheating could be obtained (13.0%), providing the first strong evidence for the validity of this questioning technique. This finding suggests that the application of the CWM will indeed allow researchers to obtain prevalence estimates that are unbiased by social desirability. We attribute the positive results obtained for the CWM in this study and previous investigations (e.g., Jann et al., 2012; Kundt et al., 2013) to its ease of use (Yu et al., 2008), comparably high comprehensibility, and perceived privacy protection (Hoffmann et al., 2015). Compared to competing approaches as, for example, the SLD (Moshagen et al., 2012), the CWM does not provide an estimate of the proportion of respondents non-adherent to the instructions. This potential disadvantage of the model might, however, be compensated for by the feature of response symmetry (cf. Ostapczuk et al., 2009): If participants are declined the opportunity to resort to a "safe" answer option, cheating detection might, indeed, be expendable.

Previous studies implementing similar anagram tasks have reported even higher cheating rates of 23% (Halevy et al., 2013) to 73% (Gino & Mogilner, 2014). The lower prevalence of cheating behavior in our sample may have been due to differences in the format of the task. In contrast to the original word-jumble task (Wiltermuth, 2011), our anagram-cheating task consisted of three items, and participants were asked merely to report the number of anagrams they had solved regardless of whether the anagrams were solved consecutively. Furthermore, our target words were presented just before participants reported their personal performance. Participants were thus well aware of the performance they had just given. Consequently, respondents had to lie rather blatantly to claim that they had solved all three anagrams. Another possible explanation for the somewhat lower cheating rate we observed may lie in the reward respondents were offered in the present study. In Wiltermuth (2011), Halevy et al. (2013), and Gino and Mogilner (2014), participants were guaranteed a fixed amount of money for each anagram they solved, and this money was presumably paid out directly after the experiment. In the present study, a better performance allowed participants only to participate in a gift-certificate lottery, which might have had a weaker effect on their motivation to cheat. Finally, differences in cheating rates may also be explained by differences in the samples and the survey setting.

In conclusion, we developed and employed a new online anagram-cheating task that can be used to induce a socially undesirable attribute with a known prevalence. This task is simple, efficient, and can easily be adopted, offering researchers the potential to conduct further strong validation studies with a reasonable amount of effort. Future studies investigating sensitive attributes and the influence of socially desirable responding might therefore profit from employing this online anagram task. Most importantly, however, the present investigation is the first to provide strong evidence that the CWM is convincingly capable of obtaining valid prevalence estimates of sensitive attitudes and behaviors. We therefore conclude that the CWM appears to be a very promising indirect questioning technique that can be used to successfully control for social desirability on surveys of sensitive behavior.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at http://dx.doi.org/10.1027/1618-3169/a000304

ESM 1. Data Pilot study 1 (Excel file).
This Excel data sheet contains the raw data collected for pilot study 1.
ESM 2. Data Pilot study 1 (SAV file).
This SPSS data set contains the raw data collected for pilot study 1.
ESM 3. Analysis Pilot study 1 (R file).
This R script contains the R syntax required to recompute the results reported for pilot study 1.
ESM 4. Data Pilot study 2 (Excel file).
This Excel data sheet contains the raw data collected for pilot study 2.
ESM 5. Data Pilot study 2 (SAV file).
This SPSS data set contains the raw data collected for pilot study 2.
ESM 6. Analysis Pilot study 2 (R file).
This R script contains the R syntax required to recompute the results reported for pilot study 2.
ESM 7. SPSS Data set Main study (SAV file).
This SPSS data set contains the raw data collected for the main study.

ESM 8. SPSS Syntax Main study (SPS file).
This R script contains the SPSS syntax required to recompute the answering frequencies and results reported for the main study.
ESM 9. MultiTree project Main study (mpt file).
This MultiTree project contains answering frequencies and equations to gain parameter estimates as reported for the main study.

References

Batchelder, W. H. (1998). Multinomial processing tree models and psychological assessment. *Psychological Assessment, 10*, 331–344. doi: 10.1037/1040-3590.10.4.331

Dominowski, R. L. (1966). Anagram solving as a function of letter moves. *Journal of Verbal Learning and Verbal Behavior, 5*, 107–111. doi: 10.1016/S0022-5371(66)80002-6

Dominowski, R. L. (1967). Anagram solving as a function of bigram rank and word frequency. *Journal of Experimental Psychology, 75*, 299–306. doi: 10.1037/H0025060

Gilhooly, K. J., & Johnson, C. E. (1978). Effects of solution word attributes on anagram difficulty – a regression analysis. *The Quarterly Journal of Experimental Psychology, 30*, 57–70. doi: 10.1080/14640747808400654

Gino, F., & Mogilner, C. (2014). Time, money, and morality. *Psychological Science, 25*, 414–421. doi: 10.1177/0956797613506438

Halevy, R., Shalvi, S., & Verschuere, B. (2013). Being honest about dishonesty: Correlating self-reports and actual lying. *Human Communication Research, 40*, 54–72.

Hilbig, B. E., & Hessler, C. M. (2013). What lies beneath: How the distance between truth and lie drives dishonesty. *Journal of Experimental Social Psychology, 49*, 263–266. doi: 10.1016/j.jesp.2012.11.010

Hoeglinger, M., Diekmann, A., & Jann, B. (2014). *Do RRT and the Crosswise Model produce valid measurements of sensitive items? An online validation study.* Paper presented at the XVIII ISA World Congress of Sociology, July 13–19, 2014, Yokohama, Japan.

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.

Hoffmann, A., Schmidt, A. F., Waubert de Puiseau, B., & Musch, J. (2015). *On the comprehensibility and perceived privacy protection of indirect questioning techniques*. Manuscript submitted for publication.

Hu, X., & Batchelder, W. H. (1994). The statistical-analysis of general processing tree models with the Em algorithm. *Psychometrika, 59*, 21–47. doi: 10.1007/Bf02294263

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

Johnson, D. M. (1966). Solution of anagrams. *Psychological Bulletin, 66*, 371–384. doi: 10.1037/H0023886

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment, 22*, 420–432. doi: 10.1037/A0019265

Korndörfer, M., Krumpal, I., & Schmukle, S. C. (2014). Measuring and explaining tax evasion: Improving self-reports using the Crosswise Model. *Journal of Economic Psychology, 45*, 18–32. doi: 10.1016/j.joep.2014.08.001

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

Küfner, C. P., Dufner, M., & Back, M. D. (2014). Das Dreckige Dutzend und die Niederträchtigen Neun – Kurzskalen zur Erfassung von Narzissmus, Machiavellismus und Psychopathie [The Dirty Dozen and the Naughty Nine: Short scales for the assessment of narcissism, Machiavellianism, and psychopathy]. *Diagnostica, 61*, 1–16. doi: 10.1026/0012-1924/a000124

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

Lemay, E. H. (1972). Anagram solutions as a function of task variables and solution word models. *Journal of Experimental Psychology, 92*, 65–68. doi: 10.1037/H0032164

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

Mayzner, M. S., & Tresselt, M. E. (1958). Anagram solution times: A function of letter order and word-frequency. *Journal of Experimental Psychology, 56*, 376–379. doi: 10.1037/H0041542

Mayzner, M. S., & Tresselt, M. E. (1959). Anagram solution times: A function of transition-probabilities. *Journal of Psychology, 47*, 117–125.

Mendelsohn, G. A. (1976). An hypothesis approach to the solution of anagrams. *Memory & Cognition, 4*, 637–642. doi: 10.3758/Bf03213228

Mendelsohn, G. A., & Obrien, A. T. (1974). The solution of anagrams – a reexamination of the effects of transition letter probabilities, letter moves, and word frequency on anagram difficulty. *Memory & Cognition, 2*, 566–574. doi: 10.3758/Bf03196922

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*, 42–54.

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

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*, 1–6.

Nederhof, A. J. (1985). Methods of coping with social desirability bias – a review. *European Journal of Social Psychology, 15*, 263–280. doi: 10.1002/ejsp.2420150303

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. (2011). Estimating the prevalence of negative attitudes towards people with disability: A comparison of direct questioning, projective questioning and randomised response. *Disability and Rehabilitation, 33*, 1–13. doi: 10.3109/09638288.2010.492067

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological*

attitudes (Vol. 1, pp. 17–59). San Diego, CA: Academic Press.
Phillips, D. L., & Clancy, K. J. (1972). Some effects of social desirability in survey studies. *American Journal of Sociology, 77*, 921–940. doi: 10.1086/225231
Quasthoff, U., Richter, M., & Biemann, C. (2006). *Corpus portal for search in monolingual corpora*. Paper presented at the fifth international conference on Language Resources and Evaluation, LREC, Genoa.
Questback. (2014). *Unipark EFS Survey 10.2*. Retrieved from http://www.unipark.de
Read, T. R., & Cressie, N. A. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York, NY: Springer.
Ruedy, N. E., Moore, C., Gino, F., & Schweitzer, M. E. (2013). The cheater's high: The unexpected affective benefits of unethical behavior. *Journal of Personality and Social Psychology, 105*, 531–548. doi: 10.1037/A0034231
Shalvi, S., Dana, J., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes, 115*, 181–190. doi: 10.1016/j.obhdp.2011.02.001
Terakoa, T. (1959). Effects of letter-orders and material words on the anagram solution. *Japanese Journal of Psychology, 30*, 253–263.
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*, 623–641. 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. (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.
Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes, 115*, 157–168. doi: 10.1016/j.obhdp.2010.10.001
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
Zipf, G. K. (1935). *The psycho-biology of language*. Oxford, UK: Houghton-Mifflin.

Received October 20, 2014
Revision received June 2, 2015
Accepted June 2, 2015
Published online December 21, 2015

Adrian Hoffmann

Department of Experimental Psychology
University of Düsseldorf
Universitätsstrasse 1, Building 23.03, Floor 00, Room 27
40225 Düsseldorf
Germany
Tel. +49 211 81-12064
Fax +49 211 811-1753
E-mail adrian.hoffmann@uni-duesseldorf.de

Appendix A

Construction of the Anagram-Cheating Task

To identify anagrams with very high versus very low difficulties, we made use of three anagram characteristics that had been found to be associated with item difficulty in previous studies. First, anagrams with frequently used target words are usually easier to solve than anagrams with rarely used target words (e.g., Dominowski, 1967; Lemay, 1972; Mayzner & Tresselt, 1958, 1959; Mendelsohn, 1976). Second, when a larger number of moves are needed to transform the anagram letters into the target word, the difficulty of an anagram usually increases (e.g., Dominowski, 1966; Mendelsohn & Obrien, 1974). Finally, anagrams with a high initial similarity to their respective target word are presumably easier to solve. Previous studies have measured this similarity using the number of letters in the anagram that remain in the correct sequence of the target word (Gilhooly & Johnson, 1978, p. 61) or by considering the correlations between the letter sequences of the anagram and the target word (e.g., Johnson, 1966; Terakoa, 1959). We considered all three of these variables while compiling our preliminary item pool. As a basis for the identification of potential target words, we used the German corpus provided by the *Deutscher Wortschatz* project (Quasthoff, Richter, & Biemann, 2006). This corpus contains information on the frequency classes of over 1,000,000 words in the German language based on a random sample of 1,000,000 sentences drawn from the German Wikipedia (http://de.wikipedia.org) in 2010. In this corpus, the frequency class of a word denotes its relative frequency in relation to the most frequent word in the corpus and is computed as

$$N = 0.5 - \log_2\left(\frac{F_i}{F_{\max}}\right) \tag{3}$$

with F_i denoting the frequency of the word in question and $F_{\max}$ denoting the frequency of the most common word in the corpus (cf. Zipf, 1935). Hence, higher values indicate lower frequencies. Aside from providing linguistic corpora, the *Deutscher Wortschatz* project also offers various services, some of which were used in this study. For example, words can be checked for grammatical properties and for anagrams in particular. To avoid enabling participants to derive hypotheses about item difficulties based on word length and to allow the use of clear-cut instructions, we extracted a set of six-letter German nouns in their basic form from the corpus but did not include any personal names or proper nouns. Words that had another German word as an anagram and words with double letters were also excluded to ensure that all anagrams would have only one solution. From the remaining word pool, we chose 32 words with a high frequency in the German language (frequency classes 7–11) as possible target words for easy anagrams, and 13 with a low frequency (frequency classes 18–22) as possible target words for difficult anagrams. To create potentially easy anagrams, we jumbled the letters in a way that ensured that participants would need to move only one letter to solve it, that five letters would already be in the correct order of the target word, and that the rank correlation between the letter sequences of the target word and the anagram would be high ($\tau > .50$). Difficult anagrams were constructed by jumbling the letters in a way that ensured that the maximum possible number of five letter moves would be needed to find the solution, that none of the letters would remain in the correct order of the target word, and that the rank correlation between the letter sequences of the target word and the anagram would be close to zero ($-.10 < \tau < .10$). A full list of all 45 anagrams thus identified for our first pilot study, along with their respective target words and item characteristics, can be obtained from Table A1.

Table A1. All 45 anagrams presented on the first pilot study

Item	Target word (translation)	Anagram	Frequency class	Letter sequence	τ	p
Easy items (expected difficulty: $p > .99$; distance to solution: one letter move)						
1	KIRCHE (church)	CKIRHE	7	412356	0.60	0.949
2	MONTAG (Monday)	NMOTAG	7	312456	0.73	0.993
3	HERBST (autumn)	HERSTB	8	123564	0.73	0.978
4	GEFAHR (danger)	GERFAH	8	126345	0.60	0.949
5	MENSCH (human)	MENHSC	8	123645	0.60	0.993
6	PRAXIS (practice)	APRXIS	9	312456	0.73	1.000
7	URTEIL (verdict)	URLTEI	9	126345	0.60	0.934
8	SCHULD (guilt)	USCHLD	9	412356	0.60	0.941
9	HEIMAT (home)	EIMHAT	9	234156	0.60	0.971
10	SCHUTZ (protection)	SCHTZU	9	123564	0.73	0.978
11	DIENST (service)	IEDNST	9	231456	0.73	0.971
12	WUNSCH (wish)	NWUSCH	9	312456	0.73	0.971
13	FREUND (friend)	UFREND	9	412356	0.60	0.978
14	JUGEND (youth)	JUGNDE	9	123564	0.73	0.963
15	UMWELT (environment)	UMELTW	9	124563	0.60	1.000
16	TERMIN (appointment)	ERTMIN	10	231456	0.73	0.919
17	UMFANG (coverage)	UMFGAN	10	123645	0.60	0.985
18	SCHLAG (beat)	SCLAGH	10	124563	0.60	0.985
19	ABWEHR (defense)	ABWREH	10	123645	0.60	0.926
20	BEZIRK (district)	BEZKIR	10	123645	0.60	0.956
21	FLUCHT (escape)	CFLUHT	10	412356	0.60	0.993
22	JUSTIZ (justice)	SJUTIZ	10	312456	0.73	0.993
23	INHALT (content)	INHLTA	10	123564	0.73	0.993
24	GIPFEL (peak)	PGIFEL	10	312456	0.73	0.919
25	BESITZ (property)	BESTZI	10	123564	0.73	0.934
26	VERBOT (prohibition)	ERBVOT	10	234156	0.60	0.963
27	STRICH (line)	STRHIC	10	123645	0.60	0.963
28	SCHILD (shield)	ISCHLD	11	412356	0.60	0.993
29	OBJEKT (object)	OBJTEK	11	123645	0.60	0.993
30	GEHALT (salary)	EHGALT	11	231456	0.73	0.978
31	SYMBOL (symbol)	BSYMOL	11	412356	0.60	0.949
32	EXPORT (export)	EXTPOR	11	126345	0.60	0.934
Difficult items (expected difficulty: $p < .01$; distance to solution: four letter moves)						
33	URINAL (urinal)	IRLAUN	18	326514	−0.07	0.022
34	ALBINO (albino)	LOANIB	18	261543	−0.07	0.088
35	GLUCKE (hen)	CUGEKL	18	431652	−0.07	0.147
36	FUNZEL (dim light)	ZULFEN	18	426153	−0.07	0.059
37	EILZUG (fast train)	LIGEUZ	19	326154	0.07	0.037
38	BARIUM (barium)	IRAMBU	19	432615	−0.07	0.007
39	KARIBU (caribou)	RAUKBI	19	326154	0.07	0.044
40	WOMBAT (wombat)	AOWTBM	19	521643	−0.07	0.206
41	METRIK (metric)	TEKIMR	20	326514	−0.07	0.015
42	TALKUM (talcum)	LTMUKA	20	316542	−0.07	0.044
43	AZIMUT (azimuth)	MIATUZ	20	431652	−0.07	0.015
44	NUKLID (nuclide)	IKUNDL	22	532164	−0.07	0.007
45	TRIOLE (triplet)	IERTLO	22	362154	−0.07	0.000

Note. τ = rank correlation between the letter sequences of the target word and the anagram; p = item difficulty observed in the first pilot study.

Appendix B

The CWM Question

Zu Ihrer Antwort von eben haben wir noch eine Frage. Wir möchten mit dieser Umfrage nämlich herausfinden, wie viele Anagramme unsere Teilnehmer durchschnittlich lösen können. Dabei ist uns wichtig, dass sich alle Teilnehmer größte Mühe bei der Bearbeitung geben. Deshalb haben besonders gute Teilnehmer, die alle drei der vorgegebenen Anagramme lösen konnten, die Möglichkeit, am Ende der Umfrage an einer Verlosung von drei Amazon-Gutscheinen im Wert von 100€, 30€ und 20€ teilzunehmen.

Hierbei ist es möglich, dass Teilnehmerinnen und Teilnehmer mehr gelöste Anagramme angeben, als sie tatsächlich gelöst haben, um Ihre Chance auf eine Teilnahme an der Verlosung zu erhöhen. Um für zukünftige Studien abschätzen zu können, wie viele Teilnehmerinnen und Teilnehmer unwahre Angaben machen, möchten wir Sie im Folgenden zu einer diesbezüglichen Stellungnahme bitten.

Ihre Antwort auf diese Frage wird streng vertraulich behandelt und hat keinerlei Auswirkungen auf die Teilnahme an der Verlosung.

Um ehrliches Antworten zu erleichtern, wird bei dieser Frage ein neuartiges statistisches Verfahren angewendet, das die Vertraulichkeit Ihrer Aussage garantiert. Durch das Verfahren sind keine Rückschlüsse auf Ihre Person möglich, der Anteil unehrlicher Antworten kann lediglich auf Gruppenebene geschätzt werden.

Dieses Verfahren möchten wir Ihnen im Folgenden kurz erläutern:

Wir präsentieren Ihnen zwei Aussagen (A und B), von denen sich eine auf die uns interessierende Fragestellung (unehrliche Antwort bei der Abfrage gelöster Anagramme) bezieht. Die andere Aussage bezieht sich auf Ihren Geburtsmonat, welchen wir nicht kennen und auch nicht erfassen werden.

Sie sollen diese Aussagen nicht einzeln beantworten, sondern lediglich angeben, ob

- Sie beiden Aussagen oder keiner der beiden Aussagen zustimmen

  oder

- Sie genau einer Aussage (egal welcher) zustimmen.

Da wir somit nicht wissen können, welche der Aussagen Sie bejaht oder verneint haben, bleibt Ihre Antwort vertraulich.

Aus Zahlen des Statistischen Bundesamtes kennen wir die relative Geburtenhäufigkeit pro Monat. Somit können wir rechnerisch im Nachhinein den Anteil unehrlicher Antworten auf Gruppenebene, also über alle Individuen hinweg, aber nicht für den Einzelfall, schätzen.

Bitte nehmen Sie nun Stellung zu den folgenden beiden Aussagen:

Aussage A: „Ich habe angegeben, mehr Anagramme gelöst zu haben, als ich tatsächlich gelöst habe."

Aussage B: „Ich bin im November oder Dezember geboren."

○ Ich stimme beiden Aussagen oder keiner der beiden Aussagen zu.

○ Ich stimme genau einer Aussage (egal welcher) zu.

Weiter

