



The impact of modes of administration on self-reports of offending: evidence from a methodological experiment with university students

Hugo S. Gomes¹ · David P. Farrington² · Marvin D. Krohn³ · Ana Cunha¹ · Julia Jurdi¹ · Bárbara Sousa¹ · Diogo Morgado¹ · Joseph Hoft⁴ · Elizabeth Hartsell⁵ · Leigh Kassem³ · Ângela Maia¹

Accepted: 31 August 2022

© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Objectives Current knowledge about the causes of offending behavior is heavily reliant on self-reports of offending (SRO). However, methodological research on the impact of modes of administration on SRO is very scarce. Further, the existing evidence conflicts with the general knowledge about responding to sensitive questions. In this study, we aimed to test whether SRO are affected by modes of administration.

Methods We carried out a methodological experiment, with a 2 (interviewer-administered vs. self-administered surveys) × 2 (paper-and-pencil vs. computer-assisted surveys) factorial design. A total of 181 undergraduate students were randomly assigned to one of these conditions and completed the International Self-Report Delinquency 3 (ISR3) questionnaire.

Results Findings showed an increased odds of reporting offending behavior in self-administered surveys, compared to face-to-face interviews. Paper-and-pencil and computer-assisted modes resulted in comparable estimates of offending.

Conclusions This experiment provides evidence that SRO provide more accurate estimates of offending behavior using self-administered surveys.

Keywords Measurement error · Survey methodology · Modes of administration · Mode effects · Sensitive questions · Delinquency

✉ Hugo S. Gomes
hugo.santos.gomes@gmail.com

¹ Psychology Research Centre (CIPsi/UM), University of Minho, Braga, Portugal

² Institute of Criminology, Cambridge University, Cambridge, UK

³ Sociology and Criminology & Law, University of Florida, Gainesville, FL, USA

⁴ University of Lynchburg, Lynchburg, VA, USA

⁵ Sam Houston State University, Huntsville, TX, USA

Introduction

Self-reports of offending (SRO) have come a long way since their early stages in the 1950s, where only a few, minor types of delinquent behaviors were included (Thornberry & Krohn, 2000). Skepticism over the utility of these methods compelled criminologists to develop a large body of research on the validity and reliability of SRO (e.g., Farrington, 1973; Huizinga & Elliott, 1986; Jolliffe et al., 2003; Piquero et al., 2014), making self-reports one of the most widely used methods in the study of offending behavior (Gomes et al., 2018). Current knowledge about the prevalence and causes of offending, as well as risk and protective factors for juvenile delinquency, are almost exclusively reliant on the self-report methodology (Cops et al., 2016; Thornberry & Krohn, 2000). However, little is known about the impact of measurement biases, such as the ones caused by modes of administration and questionnaire format, on the reported rates of offending and data quality.

In a recent systematic review of methodological experiments using SRO, Gomes et al. (2019) found 21 experiments that explored a total of 18 different manipulations of potential biases relating to modes of administration, procedures of data collection, and questionnaire design. In this study, contrary to the large body of research on sensitive questions (e.g., Gnambs & Kaspar, 2015; Richman et al., 1999; Tourangeau & Yan, 2007), the methodological experiments on SRO failed to show any evidence of the benefits of self-administration over face-to-face interviews. The lack of evidence for mode effects on SRO led influential studies on crime measurement to conclude that self-reports are valid and stable over different modes of administration (e.g., Thornberry & Krohn, 2000). However, offending behavior is a highly sensitive topic (Gomes et al., 2022), and unless there are specific features of criminal behavior, the disclosure of offending should be subject to mode effects, at least to the same extent as other types of sensitive behaviors.

In the case that SRO are, in fact, affected by modes of administration, the failure to identify these mode effects will lead researchers to apply unstandardized measurement methods, resulting in biased outcomes and, ultimately, misleading conclusions about offending behavior. In the present study, we have developed a methodological experiment carried out in Portugal with a 2 (modes of administration: interviewer-administered vs. self-administered surveys) \times 2 (modes of data collection: paper-and-pencil vs. computer-assisted surveys) factorial design, in order to test whether or not SROs are affected by modes of administration.

Sensitive questions

Sensitive topics in survey research can be defined as intrusive, posing a threat of disclosure, and eliciting socially desirable answers (Tourangeau & Yan, 2007; Tourangeau et al., 2000). An intrusive question can be construed as an inappropriate invasion of privacy. In this sense, the question itself is intrusive, independently of the participant's truthful response. The dimensions of threat of disclosure and social desirability, on the other hand, are a product of the participant's past experience and the perceived likelihood of their answers becoming known to other parties. A question on bicycle theft, for example, is nonconsequential for participants who

have never committed such behavior, even if their answers were to become known to other people outside the research study. Participants who have stolen a bicycle, on the other hand, may experience feelings of shame, guilt, or fear of criminal consequences and thus refrain from providing a truthful answer to this question. As a result, respondents to sensitive questions may tend to systematically underreport their socially undesirable behavior (Tourangeau et al., 2000).

Evidence for the tendency to underreport sensitive behavior is well documented in the literature. For example, Liber and Warner (2018) compared data from cigarette-tax collections and nationwide surveys and concluded that respondents consistently underreport cigarette consumption over time. Giguère et al. (2019) used biomarkers of recent semen exposure among female sex workers in early antiretroviral treatments and concluded that respondents often underreport unprotected sexual intercourse. Studies using biomarkers to determine substance use (provided from blood, urine, saliva, or hair samples) show that respondents consistently underestimate consumption, such as alcohol (e.g., Kabashi et al., 2019; Littlefield et al., 2017; Vinikoor et al., 2018) and other drugs (e.g., Gerdtz et al., 2020; Palamar et al., 2021). Clark and Tiffit (1966) used the polygraph as an external criterion for SRO and found evidence of underreporting of deviant behaviors. Further, studies using indirect measures consistently result in higher rates of reporting sensitive behavior than in direct questioning (Druckman et al., 2015; Kirtadze et al., 2018), including reports of offending behavior (e.g., Wolter & Laier, 2014). Because respondents to sensitive questions tend to underreport their socially undesirable behavior, survey researchers have explored methods to overcome the effects of question sensitivity. For example, measurement methods that provide anonymity and confidentiality to a respondent consistently result in higher rates of sensitive behavior (Bradburn et al., 2004).

The systematic bias of reporting higher rates of sensitive behavior in less threatening measurement conditions, where the motivation to provide socially desirable answers is reduced, cannot be explained by chance, memory faults, or the usual reporting error in survey bias (e.g., Schwarz, 1999). Rather, this evidence is consistent with the deliberate misreporting hypothesis (Bradburn et al., 1979; Tourangeau et al., 2000). According to the idea of deliberate distortion, respondents to sensitive questions deliberately edit their answers in order to avoid the embarrassment or consequences of admitting such behaviors. As a consequence, survey researchers have created the “more is better” assumption, in which measurement conditions that result in higher estimates of a socially undesirable behavior are assumed to be the most accurate (Tourangeau & Yan, 2007). This assumption is especially useful in behaviors where there is no gold standard to which self-reported information can be compared, such as offending behavior.

Modes of administration

One key variable that has repeatedly been shown to affect participants' disclosure of sensitive behavior is modes of administration (Richman et al., 1999; Tourangeau & Yan, 2007). Mainly, self-administration of a questionnaire, in contrast to interviewer-administered modes, results in a steep effect in increasing participants' willingness to report sensitive behavior (Sudman & Bradburn, 1974). In face-to-face interviews, participants

are requested to disclose their sensitive behavior to a third person (i.e., the interviewer). This is expected to affect participants' perceptions of confidentiality and anonymity, as well as social desirability, causing the above-described mode effects (Schwarz et al., 1991). Methodological experiments have provided evidence that self-administration causes increased rates of reporting multiple types of sensitive behavior, such as undesirable academic attributes (Kreuter et al. 2008), disclosure of non-heterosexual identity (Robertson et al. 2018), number of sexual partners (Jobe et al., 1997), suicidal ideation (Lee et al. 2019), and drug use (e.g., Aquilino, 1994; Butler et al., 2009; Schober et al., 1992; Turner et al., 1992). Tourangeau and Yan (2007) reviewed the survey methodological research on sensitive topics and concluded that respondents are more likely to disclose socially undesirable behaviors in self-administered conditions. Further, Tourangeau and Yan (in press) found that self-administration, in comparison to face-to-face interviews, resulted in an increase of reports of illicit drug use by 30%.

Survey research is increasingly transitioning from traditional paper-and-pencil questionnaires to computer-assisted modes of data collection. Computerized surveys are cheaper, they eliminate the need for printed questionnaires, data are automatically stored in databases and thus reduce data entry error, and computers allow for more complex branching questionnaires with skip questions, etc. (Lucia et al., 2007). Additionally, authors have suggested that computer-assisted modes increase perceived anonymity (e.g., Trau et al., 2013), raising the question of whether computerized modes of data collection impact participants' willingness to disclose sensitive behavior. The research on this particular question is fairly inconsistent. Some researchers have found no evidence of mode effects caused by modes of data collection (e.g., Bates & Cox, 2008; Knapp & Kirk, 2003). Further, the meta-analysis carried out by Dodou and de Winter (2014) found no differences in social desirability between paper-and-pencil and computer-assisted surveys.

On the other hand, some methodological experiments have found higher rates of disclosure in paper-and-pencil questionnaires (e.g., Beebe et al., 1998, 2006), while others have found results in the opposite direction, indicating higher reports of sensitive behavior in computer-assisted modes (e.g., Brener et al., 2006). Richman et al. (1999) carried out a meta-analysis and found 61 experiments comparing results obtained in computer-assisted and paper-and-pencil questionnaires (a total of 673 effect sizes). They concluded that, within self-administered modes, computer-assisted surveys resulted in a higher prevalence of sensitive behavior disclosure. More recently, Gnambs and Kaspar (2015) focused on methodological experiments comparing self-administered disclosure in paper-and-pencil and computer-assisted modes of data collection (39 studies and 460 effect sizes). These authors found that computer-assisted surveys resulted in an increased odds of reporting sensitive behavior, especially for highly sensitive topics.

The impact of modes of administration on self-reports of offending

Criminal behavior is a highly sensitive topic. Offenders naturally try to conceal their illegal behavior, and they may feel ashamed or regret their delinquent practices. The disclosure of offending behavior not only causes embarrassment and socially desirable answers, but offenders may also fear potential criminal consequences

(Thornberry & Krohn, 2000). Gomes et al. (2022) developed an assessment of question sensitivity based on the three-dimensional definition proposed by Tourangeau and Yan (2007). These authors showed that most offending questions scored higher on topic sensitivity than a question about sexual behavior, especially the more serious and violent offenses which participants rated as very highly sensitive (Gomes et al., 2022). For all these reasons, SRO are expected to be subject to reporting bias, at least to the same extent as other types of sensitive questions.

Unfortunately, methodological research on the response biases of SRO is very scarce. Gomes et al. (2019) systematically reviewed methodological experiments exploring potential response biases in the collection of SRO. In this review, the comparison between self-administered surveys using paper-and-pencil and computer-assisted modes of data collection was the most replicated manipulation within the SRO methodological literature ($k=10$). Results were very inconsistent. Five experiments found evidence showing higher reports of offending in paper-and-pencil conditions, while the other five experiments showed higher disclosure in computer-assisted modes. However, similar to previous reviews (Gnambs & Kaspar, 2015; Richman et al., 1999), the overall effect of modes of data collection on SRO showed that computer-assisted modes resulted in higher rates of reporting of sensitive behaviors, though this was only marginally significant.

As for the impact of modes of administration on SRO, Gomes et al. (2019) found three studies that carried out a total of four experimental comparisons testing the effect of self-administration on respondents' disclosure of offending behavior. Three experiments compared face-to-face interviews with paper-and-pencil questionnaires, and one of these studies also included a comparison between face-to-face interviews and audio-computer-assisted self-interview (ACASI). Results showed no significant effect of self-administration on participants' rates of reported offenses. These results disagree with the general evidence regarding self-reports of sensitive behavior (e.g., Tourangeau & Yan, 2007). However, it is worth considering that two of these studies were carried out more than 40 years ago (i.e., Hindelang et al., 1981; Krohn et al., 1974), and the third study was developed with the objective of testing mode effects on reports of risky behavior, and only two types of offenses (i.e., carrying a weapon/gun and engaging in abusive/violent behavior after drinking) were included (Potdar & Koenig, 2005). These features may have limited the ability of these studies to find evidence of mode effects, and relying solely on these findings to conclude that SRO are not affected by modes of administration may be misleading. In sum, the question about what are the best practices to measure SRO is far from settled, and more methodological research using contemporary questionnaires of offending behavior is needed.

The present study

The aim of this study was to test whether SRO are affected by modes of administration and modes of data collection. The lack of evidence showing mode effects on SRO led influential reviews of crime measurement to conclude that modes of administration did not affect participants' willingness to report offending behavior

(e.g., Thornberry & Krohn, 2000). However, if the disclosure of criminal behavior is, indeed, affected by modes of administration, similarly to the disclosure of other types of sensitive topics, then using unstandardized modes of administration may have resulted in biased conclusions about criminal behavior. Further, with the progressive transition into computerized modes of data collection, it is important to test the extent to which computer-assisted modes affect participants' reports of offending behavior in comparison to the traditional paper-and-pencil questionnaires.

In order to assess the impact of modes of administration and modes of data collection on SRO, we conducted a methodological experiment. This experiment followed a 2 (modes of administration: interviewer-administered vs. self-administered surveys) \times 2 (modes of data collection: paper-and-pencil vs. computer-assisted surveys) factorial design in which participants were randomly assigned to one of the experimental conditions. Based on the findings in the literature about sensitive topics, we predicted that participants in the self-administered modes would report higher rates of offending behavior than participants in face-to-face interviews (Hypothesis 1) and that participants in computer-assisted modes of data collection would report higher rates of offending compared to participants assigned to the paper-and-pencil modes (Hypothesis 2).

Methods

Participants

One hundred and eighty-one students from a large University in the North of Portugal, mostly female (90.6%, $n=164$), aged between 18 and 50 years ($M=20.57$, $SD=3.66$), participated in this experiment in exchange for course credits.

Design

The present study followed a 2 (modes of administration: interviewer-administered vs. self-administered surveys) \times 2 (modes of data collection: paper-and-pencil vs. computer-assisted surveys) experimental design. The crossing of these manipulations resulted in four experimental conditions: paper-and-pencil interviewer-administered interviews (PAPI); computer-assisted interviewer-administered interviews (CAPI); paper-and-pencil self-administered questionnaires (SAQ); and computer-assisted self-administered questionnaires (CASI). Participants were randomly assigned to one of these survey methods and completed the same questionnaire.

Instruments

Participants in this study completed a questionnaire composed of three main sections. First, we have included a section on socio-demographic information (e.g., sex, age, education, and income). In the second section, participants were asked to complete questions about multiple sensitive behaviors, which included the offending

behavior questionnaire. Both the socio-demographic section and the offending behavior questionnaire were drawn from the International Self-Report Delinquency 3 questionnaire (ISRD3; Enzmann et al., 2018; Portuguese version by Martins et al., 2015).

Behavioral questions followed the layout set by the ISRD3 questionnaire, in which questions were asked referring to lifetime prevalence and, in case of positive responses, participants were referred to an open-ended follow-up question about past-year incidence. Similar to previous ISRD3 studies (e.g., Doelman et al., 2021), the offending indexes were based on 12 questions on different types of deviant behavior (i.e., vandalism, shoplifting, burglary, bicycle theft, car theft, stealing from a car, stealing from a person, carrying a weapon, robbery, group fight, assault, and drug sales). We have considered lifetime and past-year prevalence rates in order to create two SRO indexes based on the variety of offending (Sweeten, 2012). Also, we have divided these offenses into lifetime and past-year composite variables of offending based on two levels of offending seriousness, i.e., property offenses (vandalism, shoplifting, burglary, and stealing from someone or a vehicle) and violent offenses (group fights, carrying a weapon, robbery, and assault) (Doelman et al., 2021).

In the third section of our questionnaire, we included measures of social desirability and participants' perceptions of privacy and anonymity. Social desirability was assessed using the Socially Desirable Response Set 5 (SDRS-5; Hays et al., 1989; Portuguese version by Pechorro et al., 2016). This is a five-item brief questionnaire (e.g., "I am always courteous even to people who are disagreeable"). Participants' perceptions of privacy and anonymity regarding their participation in this study were assessed using two ancillary questions ("I wish I could have taken the survey in a more private place" and "I am confident that the answers I gave in this survey will never be linked with my name", respectively) developed by Denniston et al. (2010). Independently of the experimental condition, all participants completed the third section of this questionnaire in a self-administered mode in order to reduce potential social desirability effects.

Procedure

Participants were recruited through the platform of exchanging course credits for participation in psychological experiments. Further, the researcher made a presentation at the end of several classes in order to recruit more participants to participate in exchange for course credits. Participants enrolled in the experiment through a doodle calendar and met the researcher in a classroom. Students were randomly assigned to one of the four experimental conditions and completed the experiment individually in a classroom in the sole presence of the researcher. Ethical approval for this experiment was provided by the Portuguese university's Institutional Review Board. Data collection was carried out from March 2018 to May 2019.

In the classroom, the researcher obtained informed consents from the participants and explained that we were interested in studying how people responded to questionnaires about sensitive topics, that they would be answering questions on personal

experiences such as offending, and that their participation in this experiment would take about 30 min. The researcher also stated that students' answers were anonymous and that their participation was confidential and voluntary. Respondents who were interested in participating in the experiment signed the informed consent, which was archived next to others in order to ensure the anonymity of participants.

Students were then randomly assigned to one of the four possible experimental conditions (i.e., PAPI, CAPI, SAQ, and CASI). In the personal interview conditions, the interviewer read the questions appearing either on the questionnaire (i.e., PAPI) or on a computer screen (i.e., CAPI) to the participants, and the interviewer ticked/entered the response provided by the participants. Interviews were carried out by five researchers (three females) which were randomly distributed to the participants. In the self-administered conditions, after providing the instructions, the researcher would step back and the exact same questions appeared either on a questionnaire (i.e., SAQ) or on a computer screen (i.e., CASI), and participants completed the survey on their own. The computer-assisted conditions were carried out using Qualtrics software with the same questions as in the paper-and-pencil conditions.

Data analysis was developed using descriptive statistics, logistic regression models to test the impact of modes of administration and modes of data collection on offending prevalence, and negative binomial regression models to test the impact of mode effects on offending variety. For logistic regression models, effect sizes are reported as odds ratios (OR: 1.68 = small, 3.47 = medium, 6.71 = large; Chen et al., 2010). As for negative binomial regression models, effect sizes are portrayed as incidence rate ratios (IRR: 1.44 = small, 2.48 = medium, 4.27 = large effect size; Borenstein et al. 2009; Roos et al., 2019). Taking into consideration that our hypotheses suggested relationships in one specific direction (e.g., higher reports of offending in self-administered conditions), all statistical analyses were carried out using one-tailed tests. All statistical analyses were carried out using SPSS software.

In order to provide an alternative to the classical significance tests, we have carried out Bayesian statistical analysis. Bayesian analysis provides a further examination of our results, indicating whether or not there is evidence of the absence of mode effects while circumventing eventual limitations of not detecting mode effects when they actually exist (e.g., an underpowered study to detect small effects) (Dienes, 2014). When comparing the support for the alternative hypothesis compared to the null hypothesis (BF_{10}) and vice-versa (BF_{01}), we have considered the following rules of thumb: < 1 no evidence; 1–3 anecdotal evidence; and 3–10 moderate evidence (Lee & Wagenmakers, 2014). Bayesian analyses were carried out using JASP software.

Results

Descriptive analysis

Participants in this study were randomly assigned either to a face-to-face interview or to a self-administered survey condition, as well as either to a paper-and-pencil or to a computer-assisted mode of data collection. As illustrated in Table 1, the random allocation

Table 1 Demographic characteristics by experimental manipulations

	Modes of administration			Modes of data collection			
	Interviewer-administered	Self-administered	Test	Computer-assisted	Paper-and-pencil	Test	<i>p</i>
No. of cases	100	81		87	94		
Age (M [SD])	20.27 (2.24)	20.96 (4.93)	$t_{(170)} = -1.23$	20.64 (4.17)	20.49 (3.09)	$t_{(170)} = 0.27$.790
Sex (%)							
Female	91.0	90.1	$\chi^2_{(1)} = 0.04$	92.0	89.4	$\chi^2_{(1)} = 0.36$.550
Interviewers' sex (%)							
Female	41.4	50.0	$\chi^2_{(1)} = 1.32$	47.7	43.0	$\chi^2_{(1)} = 0.39$.531
Economic status (%)							
Worse off	1.0	3.7	$\chi^2_{(2)} = 4.03$	3.4	1.1	$\chi^2_{(2)} = 1.32$.517
Equal	94.0	85.2		89.7	90.4		
Better off	5.0	11.1		6.9	8.5		
University grade (%)							
1 st year	30.3	30.9	$\chi^2_{(3)} = 1.97$	35.6	25.8	$\chi^2_{(3)} = 4.22$.238
2 nd year	22.2	17.3		17.2	22.6		
3 rd year	32.3	40.7		31.0	40.9		
4 th year	15.2	11.1		16.1	10.8		
Social desirability							
SDRS-5 (M [SD])	1.43 (1.17)	1.35 (1.25)	$t_{(178)} = 0.44$	1.36 (1.20)	1.43 (1.21)	$t_{(178)} = -0.41$.682
Privacy (%)							
"I wish I could have taken the survey in a more private place"							
Strongly agree/agree	6.0	3.8	$\chi^2_{(1)} = 0.47$	8.0	2.2	$\chi^2_{(1)} = 3.29$.091
Anonymity (%)							
"I am confident that the answers I gave in this survey will never be linked with my name"							
Strongly agree/agree	92.0	85.0	$\chi^2_{(1)} = 2.21$	87.4	90.3	$\chi^2_{(1)} = 0.40$.527

of participants within these experimental manipulations resulted in similar demographic characteristics. No statistically significant differences were found between these manipulations for participants' age and sex, interviewers' sex, economic status, and university class year. Further, the manipulation of both modes of administration and modes of data collection did not cause any significant effect on social desirability (Table 1). As for the ancillary questions, despite a larger prevalence of respondents in computer-assisted modes wished that they had taken this survey in a more private place (compared to paper-and-pencil modes), as well as a larger prevalence of participants in face-to-face interviews reported being confident about the anonymity of this study (compared to self-administered modes), the manipulations in this experiment had no statistically significant effect on participants' perceived privacy and anonymity.

Regarding general descriptives of offending, 32.6% of participants ($n=59$) reported committing at least one type of offense during their life-course, while 12.7% ($n=23$) of our sample reported offending in the past year. Regarding offending variety, the present sample showed a mean number of types of lifetime offending of 0.57 ($SD=0.97$, $min=0$, $max=5$). Male participants reported higher offending variety ($M=1.29$, $SD=1.49$) than females ($M=0.49$, $SD=0.88$). These differences were statistically significant ($t_{(17.16)}=-2.18$, $p<0.05$).

Modes of administration (Interviewer-administered vs. self-administered surveys)

Table 2 illustrates the effect of modes of administration on participants' reports of offending behavior. Results for overall lifetime offending prevalence show that 29% of participants in the face-to-face interview condition reported at least one type of offending behavior during their lifetime, compared to a total of 37% prevalence of offenders in self-administered surveys. However, despite this difference between the two groups, the results were not statistically significant (OR=1.44, 90% CI [0.853, 2.432]). When considering types of offending separately, small effect sizes were detected both for property and violent offenses (i.e., OR > 1.68). Present findings demonstrate that self-administration of the questionnaire resulted in a statistically significant increase in the prevalence of property offenses (OR=1.78, 90% CI [1.013, 3.121]). Results for violent offenses followed a similar trend, where participants in self-administered modes (18.5%) reported a higher prevalence of offending compared to participants in personal interview conditions (11.0%), though not reaching statistical significance (OR=1.84, 90% CI [0.908, 3.723]).

Despite the low prevalence of past-year offending, our results identified small effect sizes, suggesting that self-administration had a small effect on participants' reports of overall (OR=2.61, 90% CI [1.213, 5.630]), property (OR=3.06, 90% CI [0.956, 9.787]), and violent (OR=2.27, 90% CI [0.785, 6.565]) offending, though the regression models only reaching statistical significance for overall offending. In the case of property offenses, participants in self-administered survey conditions were 3.06 times more likely to report offending in the past year than respondents in face-to-face interviews.

As for the offending variety, modes of administration caused an increased likelihood of reports of lifetime offending (Table 2). The incidence rate of overall offending

in self-administered modes of administration was significantly higher than in interviewer-administered modes (IRR=1.66, 90% CI [1.099, 2.494]). These results were also found to be statistically significant with small effect sizes for both property (IRR=1.70, 90% CI [1.024, 2.813]) and violent offenses (IRR=1.99, 90% CI [1.062, 3.746]) over the life-course. Results for past-year offending followed the same pattern of higher disclosure of offending behavior in the self-administered conditions compared to interviewer-administered conditions (Table 2). These differences were statistically significant for past-year overall offending (IRR=2.74, 90% CI [1.356, 5.550]) and property offenses (IRR=3.29, 90% CI [1.052, 10.298]). The results for past-year diversity of violent offenses followed the same pattern detecting a small effect size, although with very low offending scores, and, thus, results did not reach statistical significance (IRR=2.47, 90% CI [0.875, 6.964]).

Modes of data collection (paper-and-pencil vs. computer-assisted surveys)

The manipulation of modes of data collection showed no statistically significant impact on lifetime or past-year offending (Table 3). Regarding lifetime prevalence

Table 2 Prevalence and variety of offending by modes of administration (lifetime offending on the top; past-year offending below)

	Interviewer-administered (n=100)	Self-administered (n=81)	B	SE	p	OR/IRR	90% CI
Lifetime offending							
Prevalence (%)							
Offending (overall)	29.0	37.0	0.37	.32	.126	1.44	[0.853, 2.432]
Property offenses	21.0	32.1	0.58	.34	.046	1.78	[1.013, 3.121]
Violent offenses	11.0	18.5	0.61	.43	.077	1.84	[0.908, 3.723]
Variety (M [SD])							
Offending (overall)	0.44 (0.78)	0.73 (1.15)	0.50	.25	.022	1.66	[1.099, 2.494]
Property offenses	0.24 (0.49)	0.41 (0.69)	0.53	.31	.043	1.70	[1.024, 2.813]
Violent offenses	0.13 (0.39)	0.26 (0.61)	0.69	.38	.036	1.99	[1.062, 3.746]
Past-year offending							
Prevalence (%)							
Offending (overall)	8.0	18.5	0.96	.47	.020	2.61	[1.213, 5.630]
Property offenses	3.0	8.6	1.12	.71	.057	3.06	[0.956, 9.787]
Violent offenses	4.0	8.6	0.82	.65	.102	2.27	[0.785, 6.565]
Variety (M [SD])							
Offending (overall)	0.09 (0.32)	0.25 (0.68)	1.01	.43	.009	2.74	[1.356, 5.550]
Property offenses	0.03 (0.17)	0.10 (0.34)	1.19	.69	.043	3.29	[1.052, 10.298]
Violent offenses	0.04 (0.20)	0.10 (0.34)	0.90	.63	.076	2.47	[0.875, 6.964]

The statistical tests are logistic regression models for prevalence (i.e., OR) and negative binomial regression models for variety (i.e., IRR)

of offending, results detected a small effect size for the impact of modes of data collection on property offenses, where a larger proportion of participants reported offending behavior in paper-and-pencil modes (30.9%) than in computer-assisted modes (20.7%), though this difference was not statistically significant (OR = 1.71, 90% CI [0.968, 3.023]). Results for lifetime prevalence of violent offenses, however, showed a different trajectory with slightly higher prevalence rates of offending under computer-assisted modes, again with no statistical significance (OR = 0.91, 90% CI [0.455, 1.835]). As for past-year offending, despite slightly higher reports of overall, property, and violent offending under paper-and-pencil modes, differences were not statistically significant (Table 3). However, these analyses detected a small effect size for mode effects on the respondents' reports of past-year property offenses (OR = 2.25, 90% CI [0.704, 7.205]). Table 3 also illustrates the effects of modes of data collection on offending variety. Findings showed very similar scores of lifetime

Table 3 Prevalence and variety of offending by modes of data collection (lifetime offending on the top; past-year offending below)

	Computer-assisted (<i>n</i> = 87)	Paper-and-pencil (<i>n</i> = 94)	<i>B</i>	<i>SE</i>	<i>p</i>	OR/IRR	90% CI
Lifetime offending							
Prevalence (%)							
Offending (overall)	27.6	37.2	0.44	.32	.084	1.56	[0.918, 2.640]
Property offenses	20.7	30.9	0.54	.27	.061	1.71	[0.968, 3.023]
Violent offenses	14.9	13.8	-0.09	.42	.416	0.91	[0.455, 1.835]
Variety (<i>M</i> [<i>SD</i>])							
Offending (overall)	0.55 (1.08)	0.59 (0.87)	0.06	.25	.406	1.06	[0.706, 1.593]
Property offenses	0.28 (0.62)	0.35 (0.56)	0.24	.31	.216	1.27	[0.768, 2.108]
Violent offenses	0.22 (0.58)	0.16 (0.42)	-0.31	.38	.202	0.73	[0.394, 1.356]
Past-year offending							
Prevalence (%)							
Offending (overall)	10.3	14.9	0.42	.46	.181	1.52	[0.716, 3.211]
Property offenses	3.4	7.4	0.81	.71	.126	2.25	[0.704, 7.205]
Violent offenses	5.7	6.4	0.11	.63	.429	1.12	[0.400, 3.124]
Variety (<i>M</i> [<i>SD</i>])							
Offending (overall)	0.15 (0.60)	0.17 (0.43)	0.13	.40	.373	1.14	[0.588, 2.206]
Property offenses	0.05 (0.26)	0.07 (0.26)	0.48	.64	.227	1.62	[0.561, 4.673]
Violent offenses	0.07 (0.30)	0.06 (0.25)	-0.08	.60	.449	0.93	[0.347, 2.468]

The statistical tests are logistic regression models for prevalence (i.e., OR) and negative binomial regression models for variety (i.e., IRR)

and past-year offending variety in paper-and-pencil and computer-assisted modes of data collection.

Bayesian analysis

In order to further analyze our results, we have carried out a Bayesian ANOVA that included the two main factors under study (i.e., modes of administration and modes of data collection) explaining lifetime and past-year offending diversity. Regarding lifetime offending, Bayesian analysis showed that the model including the main effect of modes of administration was consistently the best fitting model. The comparison analysis presented anecdotal evidence that the model including modes of administration was better at explaining our results than the null model ($BF_{10} = 1.03$). On the other hand, this analysis showed moderate evidence that the null model was better at explaining our findings than the subsequent models, especially the model including the main effect of modes of data collection ($BF_{01} = 6.05$). Similar results were found for past-year offending, where the model including the main effect of modes of administration was the best model, though only anecdotally outperforming the null model ($BF_{10} = 1.11$), while the null model was moderately better at explaining our results than modes of data collection ($BF_{01} = 8.69$).

Discussion

Self-reports are the most widely used measurement method in the study of offending behavior. Subject areas such as the study of the causes of delinquent behavior are heavily reliant on this methodology, making conclusions about delinquent behavior limited by the measurement technique. However, the lack of methodological research on SRO generates doubt about the quality of self-report measures, as well as the best ways to administer questions about offending behavior. This article provides evidence from a methodological experiment with undergraduate students from a Portuguese University. In this experiment, we have tested the effects of modes of administration (i.e., face-to-face interviews vs. self-administered surveys) and modes of data collection (i.e., paper-and-pencil vs. computer-assisted surveys) on SRO. In this 2×2 factorial design experiment, participants were randomly assigned to one of the four experimental conditions and were asked to disclose whether they have committed offending behavior over their lifetime and past year.

Offending behavior is a highly sensitive topic that generates concern about socially desirable answers and poses the threat of responses being disclosed to other people outside of the study or even fear of legal repercussions. Therefore, taking into consideration the evidence available in the literature on sensitive questions, self-administration of offending questionnaires is expected to result in higher rates of self-disclosed offending behavior compared to interviewer-administered conditions, where participants are requested to disclose their offending practices to a third person. An experimental approach is required to clearly demonstrate the impact of

modes of administration on collecting such sensitive information (Tourangeau & Yan, 2007). In the present experiment, we aimed to provide evidence regarding the best practices of administering questions on offending, in order to improve the quality of SRO data.

In line with our initial hypothesis, the present results showed that participants in self-administered conditions were more likely to report offending behavior than participants in face-to-face interviews (see Fig. 1 for a summary of findings). Results showed that participants who were asked to complete the survey in a self-administered mode had a 66% increase in the rate of disclosing lifetime offending behavior compared to participants in interviewer-administered conditions. This mode effect was even higher for past-year offending, where respondents in the self-administered mode reported an increased rate of disclosing offending behavior by 2.61 times, compared to participants in interviewer-administered conditions. The evidence for the presence of mode effects found in this experiment is in line with the general literature on sensitive questions (e.g., Richman et al., 1999; Tourangeau & Yan, 2007). Requesting someone to disclose sensitive behavior to a third person, compared to completing a survey on their own, is expected to increase social desirability effects and, thus, influence participants' willingness to disclose embarrassing and criminal behavior (Bradburn et al., 1979; Tourangeau et al., 2000).

However, contrary to the literature on modes of administration, the mode effects found in this experiment were statistically significant despite the absence of differences in participants' social desirability or perception of privacy and anonymity. Social desirability was only slightly higher in face-to-face interviews, as was the wish to have taken the survey in a more private place (with no statistical significance). As for the participants' perception of anonymity, respondents in face-to-face conditions reported slightly higher confidence that their names would never be linked to their answers, also with no statistical significance. This finding seems to be contradictory to the deliberate misreporting hypothesis (Bradburn et al., 1979; Tourangeau et al., 2000), where self-administration is expected to provide greater confidence in the study's assurances about anonymity. One potential explanation for this finding may be linked to our sample. University students may be used to completing surveys and may be aware of the ethical issues involved in carrying out research and be confident that the researcher will treat their answers carefully. Nevertheless, despite similar social desirability, anonymity, and privacy throughout the manipulated modes

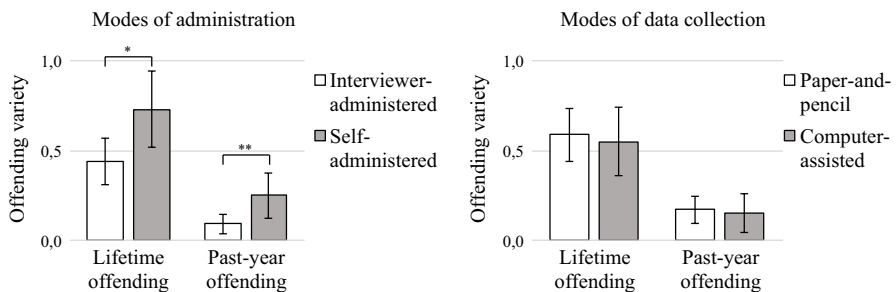


Fig. 1 The effect of modes of administration (left) and modes of data collection (right) on lifetime and past-year offending variety (error bars are 90% confidence intervals)

of administration, our results were still able to detect the presence of mode effects, in which participants in self-administered modes reported higher rates of offending than participants in face-to-face interviews. Therefore, it seems that the benefits of self-administration in improving rates of disclosing offending behavior in this study go beyond the factors of social desirability, anonymity, and privacy. More research to understand the mechanism through which self-administration causes an increased rate of reporting sensitive behavior is needed.

As for the manipulation of modes of data collection, results contrasted with our second hypothesis. According to the literature, we hypothesized that computerized modes would elicit higher rates of reporting offending behavior. However, modes of data collection showed generally no effect on reports of offending behavior, both over the life course and past year. Only in the case of property offenses, and despite the lack of statistical significance, our findings were able to detect small effect sizes in favor of higher rates of property offending in paper-and-pencil conditions, compared to computer-assisted modes. Bayesian statistical analyses provide additional support to the null results of modes of data collection, providing evidence suggestive of the lack of mode effects caused by computer vs. paper-and-pencil on respondents' willingness to disclose offending behavior. The present findings are somewhat contrary to the body of evidence from the research on sensitive topics (e.g., Gnamb & Kaspar, 2015; Richman et al., 1999), as well as from studies including offending questions (Gomes et al., 2019), where reports of sensitive behaviors are expected to be higher in computer-assisted modes. However, multiple studies have found results where behavioral reports are unaffected by modes of data collection (e.g., Baier, 2017; Hamby et al., 2006; Knapp & Kirk, 2003; Lucia et al., 2007; Trapl et al., 2013). This adds to the already inconsistent body of knowledge regarding the effects of modes of data collection on participants' willingness to provide truthful answers, and more research on the moderators of this relationship is needed.

Further, in the present experiment, participants reported very similar social desirability in both paper-and-pencil and computer-assisted modes of data collection. Also, no statistically significant differences were found for participants' perceptions about anonymity and privacy, although a somewhat higher proportion of participants in computer-assisted conditions wished that they had completed the survey in a more private place. These results are inconsistent with findings in the study of Denniston et al. (2010) that found less perceived privacy in computer-assisted modes compared to traditional paper-and-pencil modes, as well as with the study of Trau et al. (2013) that have suggested that computer-assisted modes increase participants' confidence in the study's anonymity, demonstrating once again the inconsistency of findings in the experiments comparing computer-assisted modes of data collection to the traditional paper-and-pencil questionnaires.

Limitations

Some limitations of this study need to be discussed. First, the sample in this experiment consisted of university undergraduate students. The prevalence of offending among university students is expected to be low, especially for more serious types

of offenses. The low prevalence of offending may have limited our capacity to detect mode effects because, in many cases, participants did not commit these behaviors. Also, taking into consideration that the serious offenses, as well as the most recent offenses, are regarded as the most sensitive questions (Gomes et al., 2022), the low prevalence of these serious and violent types of offenses may be a limitation to our study. Second, our samples were mostly composed of female participants. This sample characteristic may affect the generalizability of the present findings. Also, similar to the previous limitation, offending behavior is less prevalent within female participants, which may limit even more our ability to detect the impact of mode effects. Third, the current sample size ($N=181$) limits our ability to detect small mode effects, and larger samples would be preferable. However, the fact that we have detected evidence for the beneficial effects of self-administration in the reports of offending behavior in this experiment is a strong indication that SRO questionnaires are affected by mode effects. On the other hand, one might question whether this study failed to show the impact of modes of data collection due to data insensitivity (i.e., the inability to distinguish the null hypothesis from the alternative hypothesis (Dienes, 2014)). In this regard, Bayesian analysis provided additional reliability to our findings because Bayes factors allowed us to determine that present non-significant results are in support of the null hypothesis (Dienes, 2014). Further, taking into account that the support for the null model tends to increase with sample size, the Bayesian analysis gives us extra confidence that the absence of effects of modes of data collection is not due to study insensitivity (Dienes, 2014). Future studies should consider these limitations and carry out similar experiments with larger sample sizes with younger participants, from multiple backgrounds, in order to provide a larger variability of the offending variable. This would allow us to test for mode effects on more recent and more serious types of offenses, as well as to test whether the benefits of modes of administration increase with more sensitive offending questions.

Conclusions

Findings from this study showed that SRO behaviors are affected by modes of administration. Asking questions about offending behavior in self-administered conditions results in increased odds of participants' disclosure of offending behavior when compared to face-to-face interviews. Therefore, researchers using questionnaires to assess SRO should consider using self-administered modes of administration in order to increase measurement accuracy. To our knowledge, this is the first study to demonstrate the impact of self-administration on SRO, providing important information to improve the accuracy of self-report methodology to assess offending behavior and reconcile the literature on SRO and general survey research by showing how SRO, as well as other types of sensitive questions, are subject to self-administration bias (Gomes et al., 2019, 2022).

As for the effect of modes of data collection, results from this study show that asking questions using paper-and-pencil questionnaires or computer-assisted surveys resulted in mainly similar results. Further, this experiment showed that participants in paper-and-pencil and computer-assisted conditions reported similar

levels of perceived anonymity and privacy. More research on the impact of modes of data collection on SRO is needed, especially considering the gradual transition into more computerized methods and the added advantages of computer-assisted modes in reducing costs, human resources, and overcoming the limitations caused by illiteracy.

Funding The first author was supported by a doctoral grant from the Portuguese Foundation for Science and Technology (FCT—SFRH/BD/122919/2016).

Declarations

This study was conducted at the Psychology Research Centre (CIPsi/UM) School of Psychology, University of Minho, supported by the Foundation for Science and Technology (FCT) through the Portuguese State Budget (UIDB/01662/2020).

References

- Aquilino, W. S. (1994). Interview mode effects in surveys of drug and alcohol use: A field experiment. *Public Opinion Quarterly*, 58(2), 210–240. <https://doi.org/10.1086/269419>
- Baier, D. (2017). Computer-assisted versus paper-and-pencil self-report delinquency surveys: Results of an experimental study. *European Journal of Criminology*, 15(4), 385–402. <https://doi.org/10.1177/1477370817743482>
- Bates, S. C., & Cox, J. M. (2008). The impact of computer versus paper–pencil survey, and individual versus group administration, on self-reports of sensitive behaviors. *Computers in Human Behavior*, 24(3), 903–916. <https://doi.org/10.1016/j.chb.2007.02.021>
- Beebe, T. J., Harrison, P. A., Mcrae, J. A., Anderson, R. E., & Fulkerson, J. A. (1998). An evaluation of computer-assisted self-interviews in a school setting. *Public Opinion Quarterly*, 62(4), 623–632. <https://doi.org/10.1086/297863>
- Beebe, T. J., Harrison, P. A., Park, E., McRae, J. A., Jr., & Evans, J. (2006). The effects of data collection mode and disclosure on adolescent reporting of health behavior. *Social Science Computer Review*, 24(4), 476–488. <https://doi.org/10.1177/0894439306288690>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.
- Bradburn, N. M., Sudman, S., Blair, E., Locander, W., Miles, C., Singer, E., & Stocking, C. (1979). *Improving interview method and questionnaire design: Response effects to threatening questions in survey research*. Jossey-Bass.
- Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design-for market research, political polls, and social and health questionnaires*. Jossey-Bass.
- Brener, N. D., Eaton, D. K., Kann, L., Grunbaum, J. A., Gross, L. A., Kyle, T. M., & Ross, J. G. (2006). The association of survey setting and mode with self-reported health risk behaviors among high school students. *Public Opinion Quarterly*, 70(3), 354–374. <https://doi.org/10.1093/poq/nfl003>
- Buchanan, T. (2000). Potential of the internet for personality research. (In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 121–140). San Diego, CA: Academic Press.) <https://doi.org/10.1016/B978-012099980-4/50006-X>
- Butler, S. F., Villapiano, A., & Malinow, A. (2009). The effect of computer-mediated administration on self-disclosure of problems on the addiction severity index. *Journal of Addiction Medicine*, 3(4), 194–203. <https://doi.org/10.1097/ADM.0b013e3181902844>
- Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—simulation and Computation*, 39(4), 860–864. <https://doi.org/10.1080/03610911003650383>
- Clark, J. P., & Tift, L. L. (1966). Polygraph and interview validation of self-reported deviant behavior. *American Sociological Review*, 31(4), 516–523. <https://doi.org/10.2307/2090775>

- Cops, D., Boeck, A., & Pleyssier, S. (2016). School vs mail surveys: Disentangling selection and measurement effects in self-reported juvenile delinquency. *European Journal of Criminology*, 13(1), 92–110. <https://doi.org/10.1177/1477370815608883>
- Denniston, M. M., Brener, N. D., Kann, L., Eaton, D. K., McManus, T., Kyle, T. M., Roberts, A. M., Flint, K. H., & Ross, J. G. (2010). Comparison of paper-and-pencil versus Web administration of the youth risk behavior survey (YRBS): Participation, data quality, and perceived privacy and anonymity. *Computers in Human Behavior*, 26(5), 1054–1060. <https://doi.org/10.1016/j.chb.2010.03.006>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. <https://doi.org/10.3389/fpsyg.2014.00781>
- Dodou, D., & de Winter, J. C. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, 36, 487–495. <https://doi.org/10.1016/j.chb.2014.04.005>
- Doelman, E. H., Luijk, M. P., Marshall, I. H., Jongerling, J., Enzmann, D., & Steketee, M. J. (2021). The association between child maltreatment and juvenile delinquency in the context of situational action theory: Crime propensity and criminogenic exposure as mediators in a sample of European youth? *European Journal of Criminology*. <https://doi.org/10.1177/14773708211013300>
- Druckman, J. N., Gilli, M., Klar, S., & Robison, J. (2015). Measuring drug and alcohol use among college student-athletes. *Social Science Quarterly*, 96(2), 369–380. <https://doi.org/10.1111/ssqu.12135>
- Enzmann, D., Kivivuori, J., Marshall, I. H., Steketee, M., Hough, M., & Killias, M. (2018). *A global perspective on young people as offenders and victims: First results from the ISRD3 study*. Springer.
- Farrington, D. P. (1973). Self-reports of deviant behavior: Predictive and stable? *Journal of Criminal Law and Criminology*, 64(1), 99–110. <https://doi.org/10.2307/1142661>
- Gerdtz, M., Yap, C. Y., Daniel, C., Knott, J. C., Kelly, P., & Braiberg, G. (2020). Prevalence of illicit substance use among patients presenting to the emergency department with acute behavioural disturbance: Rapid point-of-care saliva screening. *Emergency Medicine Australasia*, 32(3), 473–480. <https://doi.org/10.1111/1742-6723.13441>
- Giguère, K., Béhanzin, L., Guédou, F. A., Leblond, F. A., Goma-Matsétsé, E., Zannou, D. M., Affolabi, D., Kékè, R. K., Gangbo, F., Bachabi, M., & Alary, M. (2019). Biological validation of self-reported unprotected sex and comparison of underreporting over two different recall periods among female sex workers in Benin. *Open Forum Infectious Diseases*, 6(2), 1–6. <https://doi.org/10.1093/ofid/ofz010>
- Gnambs, T., & Kaspar, K. (2015). Disclosure of sensitive behaviors across self-administered survey modes: A meta-analysis. *Behavior Research Methods*, 47(4), 1237–1259. <https://doi.org/10.3758/s13428-014-0533-4>
- Gomes, H. S., Maia, A., & Farrington, D. P. (2018). Measuring offending: Self-reports, official records, systematic observation and experimentation. *Crime Psychology Review*, 4(1), 26–44. <https://doi.org/10.1080/23744006.2018.1475455>
- Gomes, H. S., Farrington, D. P., Maia, A., & Krohn, M. D. (2019). Measurement bias in self-reports of offending: A systematic review of experiments. *Journal of Experimental Criminology*, 15(3), 313–339. <https://doi.org/10.1007/s11292-019-09379-w>
- Gomes, H. S., Farrington, D. P., Krohn, M. D., & Maia, A. (2022). How sensitive are self-reports of offending?: The impact of recall periods on question sensitivity. *International Journal of Social Research Methodology*. <https://doi.org/10.1080/13645579.2022.2077529>
- Hamby, S., Sugarman, D. B., & Boney-McCoy, S. (2006). Does questionnaire format impact reported partner violence rates?: An experimental study. *Violence and Victims*, 21(4), 507–518. <https://doi.org/10.1891/0886-6708.21.4.507>
- Hays, R. D., Hayashi, T., & Stewart, A. L. (1989). A five-item measure of socially desirable response set. *Educational and Psychological Measurement*, 49(3), 629–636. <https://doi.org/10.1177/001316448904900315>
- Hindelang, M. J., Hirschi, T., & Weis, J. G. (1981). *Measuring delinquency*. Sage.
- Huizinga, D., & Elliott, D. S. (1986). Reassessing the reliability and validity of self-report delinquency measures. *Journal of Quantitative Criminology*, 2(4), 293–327. <https://doi.org/10.1007/BF01064258>
- Jobe, J. B., Pratt, W. F., Tourangeau, R., Baldwin, A. K., & Rasinski, K. A. (1997). Effects of interview mode on sensitive questions in a fertility survey. (In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwartz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 311–329). New York: John Wiley & Sons.) <https://doi.org/10.1002/9781118490013.ch13>

- Jolliffe, D., Farrington, D. P., Hawkins, J. D., Catalano, R. F., Hill, K. G., & Kosterman, R. (2003). Predictive, concurrent, prospective and retrospective validity of self-reported delinquency. *Criminal Behaviour and Mental Health*, *13*(3), 179–197. <https://doi.org/10.1002/cbm.541>
- Kabashi, S., Vindenes, V., Bryun, E. A., Koshkina, E. A., Nadezhdin, A. V., Tetenova, E. J., Koglashkin, A. J., Petukhov, A. E., Perekhodov, S. N., Davydova, E. N., Gamboa, D., Hilberg, T., Lerdal, A., Nordby, G., Zhang, C., & Bogstrand, S. T. (2019). Harmful alcohol use among acutely ill hospitalized medical patients in Oslo and Moscow: A cross-sectional study. *Drug and Alcohol Dependence*, *204*, 107588. <https://doi.org/10.1016/j.drugalcdep.2019.107588>
- Kirtadze, I., Otiashvili, D., Tabatadze, M., Vardanashvili, I., Sturua, L., Zabransky, T., & Anthony, J. C. (2018). Republic of Georgia estimates for prevalence of drug use: Randomized response techniques suggest under-estimation. *Drug and Alcohol Dependence*, *187*, 300–304. <https://doi.org/10.1016/j.drugalcdep.2018.03.019>
- Knapp, H., & Kirk, S. A. (2003). Using pencil and paper, internet and touch-tone phones for self-administered surveys: Does methodology matter? *Computers in Human Behavior*, *19*(1), 117–134. [https://doi.org/10.1016/S0747-5632\(02\)00008-0](https://doi.org/10.1016/S0747-5632(02)00008-0)
- Krohn, M. D., Waldo, G. P., & Chiricos, T. G. (1974). Self-reported delinquency: A comparison of structured interviews and self-administered checklists. *Journal of Criminal Law and Criminology*, *65*(4), 545–553. <https://doi.org/10.2307/1142528>
- Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Liber, A. C., & Warner, K. E. (2018). Has underreporting of cigarette consumption changed over time? Estimates derived from US National Health Surveillance Systems between 1965 and 2015. *American Journal of Epidemiology*, *187*(1), 113–119. <https://doi.org/10.1093/aje/kwx196>
- Littlefield, A. K., Brown, J. L., DiClemente, R. J., Safonova, P., Sales, J. M., Rose, E. S., ... & Rassokhin, V. V. (2017). Phosphatidylethanol (PEth) as a biomarker of alcohol consumption in HIV-infected young Russian women: Comparison to self-report assessments of alcohol use. *AIDS and Behavior*, *21*(7), 1938–1949. <https://doi.org/10.1007/s10461-017-1769-7>
- Lucia, S., Herrmann, L., & Killias, M. (2007). How important are interview methods and questionnaire designs in research on self-reported juvenile delinquency? An experimental comparison of Internet vs paper-and-pencil questionnaires and different definitions of the reference period. *Journal of Experimental Criminology*, *3*(1), 39–64. <https://doi.org/10.1007/s11292-007-9025-1>
- Martins, P., Mendes, S., & Fernandez-Pacheco, G. (2015, September). *Cross-cultural adaptation and online administration of the Portuguese version of ISRD3* (Paper presented at the 15th Annual Conference of the European Society of Criminology, Porto, Portugal)
- Palamar, J. J., Salomone, A., & Keyes, K. M. (2021). Underreporting of drug use among electronic dance music party attendees. *Clinical Toxicology*, *59*(3), 185–192. <https://doi.org/10.1080/15563650.2020.1785488>
- Pechorro, P., Barroso, R., Silva, I., Marôco, J., & Gonçalves, R. A. (2016). Propriedades psicométricas da Escala de Respostas Socialmente Desejáveis-5 (SDRS-5) em jovens institucionalizados [Psychometric properties of the socially desirable response set 5 (SDRS-5) with institutionalized youth]. *Psicologia*, *30*(1), 29–36. <https://doi.org/10.17575/rpsicol.v30i1.1065>
- Piquero, A. R., Schubert, C. A., & Brame, R. (2014). Comparing official and self-report records of offending across gender and race/ethnicity in a longitudinal study of serious youthful offenders. *Journal of Research in Crime and Delinquency*, *51*(4), 526–556. <https://doi.org/10.1177/0022427813520445>
- Potdar, R., & Koenig, M. A. (2005). Does audio-CASI improve reports of risky behavior? Evidence from a randomized field trial among young urban men in India. *Studies in Family Planning*, *36*(2), 107–116. <https://doi.org/10.1111/j.1728-4465.2005.00048.x>
- Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, *84*(5), 754–775. <https://doi.org/10.1037/0021-9010.84.5.754>
- Roos, C., Stein, E., Bowen, S., & Witkiewitz, K. (2019). Individual gender and group gender composition as predictors of differential benefit from mindfulness-based relapse prevention for substance use disorders. *Mindfulness*, *10*, 1560–1567. <https://doi.org/10.1007/s12671-019-01112-y>
- Schober, S. E., Caces, M. F., Pergamit, M. R., & Branden, L. (1992). Effect of mode of administration on reporting of drug use in the National Longitudinal Survey. (In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (pp. 267–276). Washington, DC: National Institute on Drug Abuse.)

- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*(2), 93–105. <https://doi.org/10.1037/0003-066X.54.2.93>
- Schwarz, N., Strack, F., Hippler, H. J., & Bishop, G. (1991). The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology*, *5*(3), 193–212. <https://doi.org/10.1002/acp.2350050304>
- Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*. Aldine Publishing Company.
- Sweeten, G. (2012). Scaling criminal offending. *Journal of Quantitative Criminology*, *28*, 533–557. <https://doi.org/10.1007/s10940-011-9160-8>
- Thornberry, T. P. & Krohn, M. D. (2000). The self-report method for measuring delinquency and crime. (In D. Duffee, R. D. Crutchfield, S. Mastrofski, L. Mazerolle, D. McDowall, & B. Ostrom (Eds.), *Criminal justice: Measurement and analysis of crime and justice* (Vol. 4, pp. 33–84). Washington, DC: National Institute of Justice)
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883. <https://doi.org/10.1037/0033-2909.133.5.859>
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.
- Tourangeau, R., & Yan, T. (in press). Reporting issues in surveys of drug use. *Substance Use and Misuse*.
- Trapl, E. S., Taylor, H. G., Colabianchi, N., Litaker, D., & Borawski, E. A. (2013). Value of audio-enhanced handheld computers over paper surveys with adolescents. *American Journal of Health Behavior*, *37*, 62–69. <https://doi.org/10.5993/AJHB.37.1.7>
- Trau, R. N., Härtel, C. E., & Härtel, G. F. (2013). Reaching and hearing the invisible: Organizational research on invisible stigmatized groups via web surveys. *British Journal of Management*, *24*(4), 532–541. <https://doi.org/10.1111/j.1467-8551.2012.00826.x>
- Turner, C. F., Lessler, J. T., Devore, J. W. (1992). Effects of mode of administration and wording on reporting of drug use. (In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (pp. 177–219). Washington, DC: National Institute on Drug Abuse.)
- Vinikoor, M. J., Zyambo, Z., Muyoyeta, M., Chander, G., Saag, M. S., & Cropsey, K. (2018). Point-of-care urine ethyl glucuronide testing to detect alcohol use among HIV-hepatitis B virus coinfecting adults in Zambia. *AIDS and Behavior*, *22*(7), 2334–2339. <https://doi.org/10.1007/s10461-018-2030-8>
- Wolter, F., & Laier, B. (2014). The effectiveness of the item count technique in eliciting valid answers to sensitive questions An evaluation in the context of self-reported delinquency. *Survey Research Methods*, *8*(3), 153–168. <https://doi.org/10.18148/srm/2014.v8i3.5819>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Hugo S. Gomes is a Fulbright Scholar and a Doctorate Researcher at the Center for Psychology at the University of Porto. He holds a PhD in applied psychology from the University of Minho, in collaboration with the Institute of Criminology, University of Cambridge. His main research interests are in research methods, developmental and life-course criminology, and experimental criminology. He was awarded the 2020's Student Paper Award by the Division of Experimental Criminology of the American Society of Criminology (ASC).

David P. Farrington is Emeritus Professor of Psychological Criminology at the Institute of Criminology, Cambridge University. His major research interest is in developmental criminology, and he is Director of the Cambridge Study in Delinquent Development, which is a prospective longitudinal survey of over 400 London males from age 8 to age 61. In addition to 905 published journal articles and book chapters on criminological and psychological topics, he has published 132 books, monographs and government publications, and 164 shorter publications (total = 1,201).

Marvin D. Krohn is an Emeritus Professor of the Department of Sociology and Criminology & Law at the University of Florida. He is primarily interested in developmental and life course approaches to the explanation of delinquency, drug use and crime. He served as a Co-Principal Investigator on the Rochester Youth Development Study and was a co-author of the *Gangs and Delinquency in Developmental Perspective*, the winner of the ASC's 2003 Most Outstanding Contribution to Criminology Award. He is a Fellow of the American Society of Criminology. Professor Krohn was awarded the Academy of Criminal Justice Science's 2016 Outstanding Mentor Award and the Southern Criminal Justice Association's 2020 Outstanding Educator Award.

Ana Cunha master's in psychology and PhD student in Applied Psychology at CIPsi, School of Psychology at UMinho with a financed project about the impact of COVID-19 pandemic on victims of domestic violence. Research interests are domestic violence, multiple victimizations, human trafficking, and trauma informed practices. She is the (co)author of national and international publications in these fields.

Julia Jurdi is an attorney graduated from Presbyterian University Mackenzie, in São Paulo (Brazil), and master in Crime, Difference and Inequality by the Institute of Social Sciences at University of Minho, in Braga (Portugal). Main research interests are recidivism, rehabilitation and reintegration.

Bárbara Sousa is a PhD student in Applied Psychology at the School of Psychology, University of Minho, Braga, Portugal; where she develops studies focused on work-family reconciliation, professional challenges, exposure to potentially traumatic events, and mental health in the police forces. Currently, she is working in a research project supported by the Foundation for Science and Technology (FCT) that focuses on the impact of police officer's exposure to potentially traumatic events on health and family relationships. Overall, she is main research interest are first responder's exposure to potentially traumatic events, complex trauma, PTSD and traumatic events in adults. She is a co-author of national and international journals.

Diogo Morgado is a PhD student in applied psychology at the Psychology Research Center (CIPsi/UM), School of Psychology at University of Minho. Currently, he is working in a research project supported by the Foundation for Science and Technology (FCT) that focuses on the relationship between peer victimization and complex trauma in adolescence. Overall, his main research interests are peer victimization, complex trauma, PTSD and traumatic events. He is a co-author of several articles published in national and international journals.

Joseph Hoft is an Assistant Professor of Criminology at University of Lynchburg. He received his doctorate in Criminology & Law from University of Florida. His areas of expertise include criminal-psychological risk behavior, plea bargaining, and pretrial detention.

Elizabeth N. Hartsell graduated from the University of Florida with her PhD in the summer of 2022. She is currently an Assistant Professor at Sam Houston State University in the Department of Criminal Justice and Criminology. Her interests include courts, substance use and mental health in the justice system, problem-solving courts, mixed methods research, and instrumentation.

Leigh Kassem holds an M.A. in Criminal Justice & Criminology from East Tennessee State University. She was the recipient of the 2020 Student Paper Award from the Academy of Criminal Justice Sciences Drugs and Alcohol Research Section and the 2020-2021 Graduate Student Teaching Award from the University of Florida. Her professional interests include juvenile delinquency, substance use, and family dynamics from a life-course perspective. She has worked closely with the "Stop Now and Plan" program at CDS Family and Behavioral Health Services in Gainesville, Florida since 2019 and is currently serving youth and families as the agency's clinical Case Manager for SNAP Under 12 and SNAP in Schools in Alachua County.

Angela Maia PhD in Clinical Psychology and Habilitation in Applied Psychology. Vice-Director of Doctoral College, Director of the Master in Justice Psychology and coordinator of the Laboratory Victims, Offenders and the Justice System and the Trauma Psychotherapy Group. Coordinates research projects funded by the Portuguese Foundation of Science and Technology and Government Departments in the area of trauma and violence, being author of several publications. Member of the Military Stress Center of the Ministry of Defense, she is interested in understanding the impact of exposure to adversity, potentially traumatic experiences and associated factors, as well as pathways from victimization to delinquency.