



Addressing Ethnic Differences in the Validity of Self-reported Criminal Behaviour Through a Social Desirability Measure

Willemijn E. Bezemer¹ · Marise Ph. Born² · Arjen S. Leerkes^{1,3}

Accepted: 9 January 2023 / Published online: 7 February 2023
© The Author(s) 2023, corrected publication 2023

Abstract

Objectives Self-reported criminal behaviour has the potential to provide clearer insights into patterns of criminality compared to using police statistics. The risk of inaccurate responses however forms a major obstacle to its validity. This study therefore examines underreporting bias of self-reported criminal behaviour among five ethnic groups and compares different methods to facilitate the creation of valid intergroup comparisons.

Methods This study includes data from the Monitor on Youthful Delinquency (N=6,218) which was connected to police suspect registrations. To identify patterns of underreporting, we compared self-reported and police recorded crime with a social desirability measure, which was adjusted to be invariant across ethnic groups. Three different methods to correct for underreporting bias were subsequently compared; partialling out the effect of social desirability, listwise deletion, and a novel technique which we named Social Desirability based Score Replacement (SDSR).

Results The study reveals that police suspects with a high social desirability score display a low likelihood to self-report crime when they have an ethnic minority background, but not when they have a native Dutch background or when they have a moderate to low social desirability score. This finding points towards systematic differences in underreporting bias. Model outcomes are shown to be significantly impacted depending on the method that is used to address this issue.

Conclusion Neglecting to correct underreporting-bias hinders the validity of intergroup comparisons of self-reported criminal behaviour. The inclusion of a social desirability measure is therefore recommended to help identify and correct underreporting bias, particularly through the use of SDSR.

Keywords Ethnicity · Self-reported criminal behaviour · Social desirability · Underreporting · Validity

✉ Willemijn E. Bezemer
bezemer@essb.eur.nl

¹ Department of Public Administration and Sociology, Erasmus University Rotterdam, Burgemeester Oudlaan 50 T15-47, 3062 PA Rotterdam, The Netherlands

² Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands

³ Maastricht Economic and Social Research Institute on Innovation and Technology, United Nations University, Maastricht, The Netherlands

Introduction

Criminology has existed as an academic discipline for more than a hundred years. Nevertheless, our insights into patterns of criminal behaviour remain limited, because incidences of criminal behaviour are notoriously difficult to measure. Most studies on crime are therefore based on police statistics, even though police data are known not to reflect patterns of criminal behaviour in a representative manner. Some offences are more likely to be policed, such as burglary and pickpocketing, while other types are less likely to be brought to justice, such as vandalism or domestic violence (CBS 2010). In addition, the characteristics of the offenders—such as their age, gender and ethnicity—influence the likelihood of police getting involved and judicial steps being taken (Grisso and Schwartz 2000; Goff et al. 2014; Boon et al. 2019). Police statistics therefore provide a skewed image of crimes that are committed and of the type of offenders who commit crimes (Buil-Gil et al. 2021; Elffers and van der Kemp 2016; Xie and Lauritsen 2012). Moreover, the correlation between criminal behaviour and crime statistics becomes smaller with each step further into the judicial system, starting from the initial suspicion to the final phase of sentencing (Kurlychek and Johnson 2019; Sellin 1931).

The lack of reliable data on criminal behaviour particularly complicates prevailing discussions around the overrepresentation of ethnic minorities in the justice system. This is exacerbated by the fact that data on the different stages of the judicial process show that overrepresentation of ethnic minorities is far greater in the later stages compared to the earlier stages (Boon et al. 2019). In the Netherlands, as in most European countries, the incarceration rates of migrant groups are many times higher than those of the native population. The largest non-western migrant groups in the Netherlands originate from Morocco, Turkey, the Dutch Antilles, and Surinam. These first two minorities mostly consist of descendants or other relatives of the guest workers who were recruited during the 1960s and early 1970s; the latter two minorities result from the country's colonial past. (Engbersen et al. 2007). Since the 1990s, research in the Netherlands has consistently shown large degrees of overrepresentation of these groups—especially youth with a Moroccan and an Antillean background—in the different phases of the criminal justice system (Leun et al. 2010). The likelihood of a person becoming imprisoned is 12.5 times higher for a person of Moroccan descent compared to a person who is a Dutch native.¹ These discrepancies are even larger among youths aged 18–25. A peak occurs within this age range in terms of judicial contacts and also in terms of the differences between the native population and migrant groups regarding the likelihood of imprisonment. The likelihood of a youth with a Moroccan descent becoming imprisoned in the Netherlands is 20 times higher than for youth without a migrant background.² While it is still largely unclear how these numbers ought to be interpreted, these numbers are systematically used as legitimization of anti-immigrant sentiments and policies by police, politicians and civilians. It is not yet known to what extent overrepresentation is the result of systematic differences between groups in terms of

¹ In 2020, 3370 people with a Moroccan background and 8810 incarcerated people with a Dutch background were incarcerated. There were 32.5 times as many Dutch people than people with Moroccan descent living in the Netherlands in 2020. Incarceration rates retrieved from:

<https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82321NED/table?fromstatweb>. Population statistics retrieved from: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37713/table?fromstatweb>.

² In 2020, among those aged 18–25, 900 people with a Moroccan descent and 1000 people with a native Dutch descent were incarcerated. There were 22 times as many Dutch people than people with a Moroccan descent in this age range living in the Netherlands in 2020, *ibid*.

criminal behaviour, and to what extent overrepresentation is caused by ethnic profiling and biases among the police and among the people who report crimes. To answer these questions, data on criminal behaviour are required which are collected independently from the judicial system.

An example of such an alternative data source is self-reported criminal behaviour. Through this method, one can obtain information on criminal behaviour both among justice-involved offenders and among non-justice-involved offenders. Undocumented criminal activities are known as “the dark figure of crime” because it is very difficult to obtain reliable information on the scope of criminal activities that have not been discovered by the police. However, studies on self-reported criminal behaviour which are based on a representative population sample and also include the full range of criminal activities are rarely carried out. This is for several reasons. Self-reporting of illegal behaviour inherently produces validity issues, which puts serious limitations on the accuracy and generalisability of its outcomes. In addition, it is expensive and time-consuming to generate a representative sample of a population with a sufficient response rate among all social groups.

The large differences between ethnic groups in terms of judicial involvement are not reflected in studies on self-reported criminal behaviour (Piquero and Brame 2008; Farrington et al. 1996; Jolliffe et al. 2003; Piquero et al. 2000). Some of these studies find equal patterns of criminal behaviour between ethnic groups and other studies find differences between groups, but to a much smaller degree than police records would suggest (Hindelang et al. 1979; Krohn et al. 2013). It is uncertain how the stark differences between police statistics and self-reported delinquency should be interpreted. Many studies on self-reported criminal behaviour conclude that their outcomes demonstrate the bias of the judicial system since the large discrepancies between ethnic groups are not reflected in these studies. However, critics argue that this discrepancy could also be the result of an invalid measurements of self-reported crime. For example, some have argued that stigmatised groups would have a higher motivation to underreport their criminal activities (Bersani and Piquero 2017). Several studies have indeed shown that the validity of self-reported behaviour on criminal and other socially undesirable activities systematically differs between ethnic groups and tends to be lower among ethnic minorities living in Europe and in North America compared to the ethnic majority group (Hindelang et al. 1979; Kirk 2006; Lab and Allen 1984; Maxfield et al. 2000; Tracy 1987; Barger 2002; He 2015; Johnson and Vijver 2003; Uziel 2010). This finding hinders our ability to use self-reported measures to make reliable comparisons between ethnic groups regarding their criminal behaviour.

To address the validity issues of self-reported criminal behaviour, this paper focuses on the concept of social desirability as an indicator of the likelihood that respondents would be willing to disclose their past criminal activities. Studies on self-reported criminal behaviour rarely include a social desirability scale, perhaps because it is not yet clearly understood how social desirability can be used to identify untruthful responses. Moreover, it may not even be clearly understood what the concept of social desirability exactly pertains. Contrasting theories prevail about its meaning. Some academics have found social desirability to be an indicator of concealment of undesirable attitudes and behaviours, suggesting that a high degree of social desirability leads to underreporting of criminal behaviour (Fernandes and Randall 1992; Kashy and DePaulo 1996; Short et al. 2009). Others conceptualise social desirability as a personality trait which relates to a socially conventional and dependable persona (Paulhus 1981), and to obedience to authority (Crowne and Marlowe 1960). This conceptualisation is supported by studies where negative correlations have been found between a socially desirable response style and criminality (Mills and Kroner 2005; Sugarman and Hotelling 1997; Tan and Grace 2008; Twenge and Im 2007). These

findings suggest that a high degree of social desirability is associated with lower levels of criminal behaviour.

These two contrasting theories on the meaning of social desirability both fail to consider that social desirability may not have a universal meaning. Its meaning may differ depending on the cultural characteristics of the population that is studied (Randall et al. 1993). The most common social desirability measure, the Marlowe-Crowne Social Desirability Scale (MCSDS), has been designed by and for educated people from western countries. It is not yet clear whether this scale produces similar meanings for non-western people, which puts the validity of intercultural comparisons into question (Johnson and Vijver 2003).

The following research question will be addressed in this paper: *How can a social desirability measurement be used to increase the validity of data on self-reported criminal behaviour?* To answer this question, we will first need to answer two other questions: *How can social desirability be measured in a way that is comparable for people with different ethnic backgrounds?* And: *How does social desirability relate to the underreporting of criminal behaviour?* To answer these questions, data from the Monitor on Youthful Delinquency from 2010 and 2015 is used. This data contains a representative sample of Dutch youth between the ages of 12 and 23 ($N=6,218$). The dataset was subsequently enriched with policing records and municipal data from Statistics Netherlands. Other studies have mostly compared self-reported criminal behaviour to police data on an aggregate level, instead of comparing its outcomes on the same set of respondents, which complicates comparability. Connecting police records to self-reported criminal behaviour on an individual level results in a unique dataset, facilitating clearer insights into the factors that affect the likelihood to become a criminal suspect.

This paper is structured as follows. First, we examine how to arrive at a measurement of social desirability that relates to the same underlying concepts for people of different ethnic backgrounds. Second, we analyse how a respondent's level of social desirability influences the likelihood of reporting crimes, in particular among respondents who are already a criminal suspect. Finally, we compare three different analytic strategies to minimise bias emerging from the underreporting of criminal activities: (1) by controlling for social desirability in the analytical model, (2) by using listwise deletion of respondents with a certain degree of social desirability, and (3) by using a novel method that we call *Social Desirability based Score Replacement* (SDSR). This technique involves the removal of scores related to self-reported criminal behaviour among respondents with the highest level of social desirability and replacing these values through the method of multiple imputation. It allows the correction of responses with a high risk of being invalid by using the response patterns of respondents who are considered to provide valid accounts of their criminal behaviour. The methodological innovations presented in this study are aimed at facilitating the creation of new and more accurate insights into patterns of criminality which occur both within and outside the view of police. These insights can in turn contribute to elucidating the overrepresentation of ethnic minorities in the justice system.

Theory

Self-reported Criminal Behaviour

Thornberry and Krohn (2000) find the introduction of self-reported criminal behaviour to be one of the most important innovations in criminological research in the twentieth

century because of its potential to provide more accurate insights into criminal behaviour. This method was first introduced in the 1940s, and further developed in the 1950s with the incorporation of carefully designed reliability and validity mechanisms (Short et al. 2009). Nowadays, however, much less attention is being paid to critical ways of ensuring validity and reducing bias of data that are based on self-reported criminal behaviour (Junger-Tas et al. 2010; Kivivuori 2011; Verbruggen et al. 2021). This present lack of rigour may be the reason why self-reported criminal behaviour has been received with criticism and even outright rejection (e.g., Junger 1990; Lew 1997) and is not taken into serious consideration by police organisations for gaining insight into patterns of criminality (also see Maxfield and Babbie 1998). Hindelang and colleagues (1979) provide an overview of the three major shortcomings of self-reported criminal behaviour that reduce its validity. First, most studies use limited samples which do not representatively reflect the population. Second, most surveys do not collect information on all types of criminal behaviour and tend to focus on minor and common offences. Finally, respondents often do not provide valid answers due to social desirability. Resolving these three issues can minimise the risk that differences with police-registered criminality are due to sample and measurement issues instead of undue selectivity.

The Marlowe-Crowne Social Desirability Scale

The likelihood that respondents are not willing to self-report on their own negative attitudes and behaviours is a major cause of the critiques which studies on self-reported criminal behaviour have received (Hanson and Bussière 1998). The scientific debate on concealment of negative traits and behaviours mostly focuses on the topic of social desirability and its most common measurement instrument, the Marlowe-Crowne Social Desirability Scale (MCSDS). This scale was developed in 1960 by the psychologists Marlowe and Crowne in order to measure the degree of social desirability among non-clinical respondents. The developers of this scale viewed social desirability as a personality trait and a habitual response style in situations of self-evaluation. The original scale contains 33 true/false items pertaining to behaviours that are socially desirable, but also uncommon or vice versa. An example of such a question is: “Before voting I thoroughly investigate the qualifications of all the candidates”. They conceptualised that this scale measures the underlying construct of a “need for approval”. This need can be expressed in two ways. Firstly, as self-deceptive enhancement, where respondents unconsciously ascribe overly favourable characteristics to themselves. Secondly, via impression management, where respondents are aware that a certain trait does not reflect their thoughts and behaviour, but they report it to project a positive image of themselves. The scale is therefore commonly used as an indicator of the likelihood that a person consciously or unconsciously will conceal undesirable attitudes and behaviours (Paulhus 1984). Respondents with a high degree of social desirability present themselves in a positive light, independent of their true behaviour (Krumpal 2013). Respondents with a low level of social desirability are prepared to acknowledge and report aspects of themselves that go against societal norms. However, Marlowe and Crowne’s definition of social desirability is not uncontested. Behavioural scientists argue that a high score on the social desirability scale may also indicate that a person simply is not likely to display undesirable behaviour (Uziel 2010). Therefore, it is important to investigate in what way social desirability bias influences survey outcomes (Fernandes and Randall 1992).

Research on social desirability has shown that its prevalence does not only differ between persons; there are also systematic differences between social groups. People who live in countries with less welfare, with a less individualistic culture, and with more collectivism, have higher average levels of social desirability (He 2015; Johnson and Vijver 2003). In addition, the type of social desirability may vary across cultures; people from more individualistic cultures are more likely to engage in self-deceptive enhancement, while people from more collectivist cultures are more likely to engage in impression management (Lalwani et al. 2006). Systematic differences have also been found in the average level of social desirability between people of different social classes (Uziel 2010) and between men and women (Barger 2002). Research from Ross and Mirowski (1984) on social desirability among Mexican people, Mexican migrants living in the USA and Anglo-Americans living in the USA, found that the Mexican migrant group displayed higher levels of social desirability in comparison to Anglo-Americans and Mexicans living in their native country. People tend to underreport negative traits when these do not only reflect negatively on themselves but also on their family or a social group which they are part from. This applies even more to people who are vulnerable to rejection in the economic or social sphere through prejudice and discrimination (Ross and Mirowski 1984).

Beretvas et al. (2002) conducted a systematic literature study on the use of the MCSDS in research and found that it is mainly used in three ways. The most common use is to correlate the scale with the focal variables of the research study. Low correlations are taken as indicators that the data are not biased by distortions resulting from social desirability. The second method is the creation of a factor analysis containing the variables of interest and the MCSDS to verify whether the items that make up these different constructs cluster on separate dimensions. The third method is the use of listwise deletion to exclude respondents with high social desirability scores. This method is based on the assumption that scores on the scale of interest are confounded by social desirability bias. Listwise deletion has been used since the 1950s to remove “overconformers” who score extremely high on the social desirability scale (Short et al. 2009). A fourth method which was not mentioned in this paper, but which has been used regularly, is the use of the MCSDS as a control variable in order to partial the effect of social desirability out of the analysis (Leerkes et al. 2019; MacDonald et al. 2017; Paulhus 1981; Saunders 1991; Sutton and Farrall 2005). This method aims to reduce the bias of inaccurate responses by only comparing respondents with similar levels of social desirability.

Data, Measures, and Analytical Strategy

Data

This study utilises data from the Youth Delinquency Survey (YDS). This monitor is designed and managed by the Research and Documentation Centre (WODC) of the Dutch government and the data-collection is carried out by Statistics Netherlands (CBS). Every five years, a new wave takes place, having started in 2005. Its purpose is to provide insights into youth delinquency to complement information that can be derived from police records. This study uses the 2010 and 2015 waves because these instalments include a social desirability measure. This measure was unfortunately dropped from the 2020 instalment of the survey, which was therefore not included. Our dataset consists of 6,218 respondents between the ages of 12 and 23, living in 318 different Dutch municipalities. These

respondents were selected based on a randomised stratified sample of municipal records with age and ethnic background being used as selection criteria. The four largest non-western migrant groups in the Netherlands are people with a Turkish, Moroccan, Antillean, and Surinamese background. These groups were oversampled to ensure a sufficient sample size that would facilitate intergroup comparisons. Earlier research has demonstrated that these different groups significantly differ from each other in the degree that criminal behaviour is underreported (Junger 1989), in their reputation in Dutch society (Van der Leun et al. 2010), and in the level of criminal justice involvement (Jennissen and Blom 2007), which points to a need to examine each group separately. To reduce non-response, the selected participants were approached up to seven times by Statistics Netherlands, resulting in a response-rate of 68% (Weijters et al. 2016). Respondents were interviewed face-to-face, but information about criminal behaviour was anonymously collected by asking respondents to fill out questions on a laptop which was positioned out of view of the interviewer. The other survey questions were afterwards matched to the data from the laptop through an encrypted key. This dataset has the potential to overcome all the major shortcomings of studies on self-reported criminal behaviour as described by Hindelang et al. (1979): The YDS consists of a representative sample, it measures almost the full spectrum of criminal activity and it includes a social desirability measure in order to address validity issues of underreporting bias.

For the purpose of this study, the data from the YDS were enriched with police data on criminal suspects. Statistics Netherlands, who store both datasets, provided anonymised identification numbers through which the police records from each of the respondents could be connected to the YDS data without breaching the anonymity of the respondents. The dataset on police suspects is cleaned every year for three consecutive years after the filing of the data. This means that suspects are removed from the dataset when they have been mistakenly recorded. By connecting these two datasets, it becomes possible to distinguish which self-reported offenders *are* and *are not* subjected to judicial involvement.

Measures

Self-reported criminal behaviour was measured by asking whether the respondent had ever committed a certain criminal offence. Questions were posed in relation to 14 light and 13 serious offences. The serious offences are: (1) violent robbery, (2) dealing hard drugs, (3) stealing something from inside a car, (4) shoplifting with a value above 10 euros, (5) a robbery attempt, (6) selling stolen goods, (7) stealing a scooter/bike, (8) burglary, (9) stealing something from the outside of a car, (10) dealing party drugs, (11) pickpocketing, (12) causing injury using a weapon, and (13) rape. The light offences are: (14) dealing soft drugs, (15) damaging a residential building, (16) damaging a vehicle, (17) buying stolen goods, (18) carrying a weapon in a nightlife area, (19) vandalising a bus, tram, metro, or train, (20) physical violence with injury, (21) threatening to intimidate, (22) shoplifting with a value below 10 euros, (23) damaging something else, (24) changing prices in a store, (25) defacing walls, trams, or buses, (26) physical violence without injury, and (27) stealing from work/school. The WODC decided whether an offence was categorized as light or serious by using the criteria of prevalence in the sample and its highest possible punitive measure in the Dutch criminal code. Offences are considered as ‘serious’ when one can be punished with more than 48 months imprisonment *and* when less than 3% of respondents self-reported to have committed this crime in the past year. The only offences that were both rare and who could not be punished with more than 48 months imprisonment were

the vandalism offences. These offences are therefore categorized as ‘light’. The offence of assault with a weapon is categorized as serious even though it can be punished with a maximum of 25 months imprisonment. Since this offence was reported by less than 3% of respondents, the researchers assumed that respondents did not report trivial incidents here, which resulted in their decision to categorise such offences as ‘serious’ (Laan and Blom 2006, p. 279). To include information in the analyses on the type, frequency, and seriousness of the crime, we created dummy variables of the different types of crimes and created a frequency scale for all the serious and all the light offences. The resulting multitude of variables on criminal behaviour will unavoidably introduce collinearity in the analysis. For that reason, we do not interpret the individual coefficients of these variables. Instead, we focus on the consequences of the inclusion of all variables related to self-reported criminal behaviour for the estimated effects of ethnic origin on the number of police registrations.

Control variables of gender and age are included. Gender is a dichotomous variable with males serving as the reference group. Age is calculated in years and standardized in the analysis through a Z score. Ethnic background was not self-reported but derived from Statistics Netherlands data. Respondents are included with a native Dutch, Moroccan, Turkish, Surinamese, Antillean, or other migration background. An ethnic minority status is ascribed when the respondent themselves or at least one of their biological parents were born in one of these countries (Weijters et al. 2016).

For the purposes of this study, the YDS was enriched post-hoc with police data on criminal suspects. This judicial dataset contains measurements of 56 different crimes. The dataset of police suspects contains information on all registered suspects in The Netherlands who were suspected of having committed one or more offences in a given year. In order to be counted as a suspect, ‘reasonable guilt’ needs to be established, which is usually the result of victim reports and/or police investigations. The dataset related to suspected offenders is subject to change in the years after its creation. For example, when a suspect is identified for a crime that was committed in a previous year, this person will be registered in the suspect database of the year in which the offence occurred. In other cases, a registered suspect may have been unduly registered, by mistake or due to a false accusation. The police register these changes in their own systems. During the yearly data provision to Statistics Netherlands, the updated dataset of a particular year is replaced with the dataset that was provided the previous year. This process continues for three consecutive years, after which the dataset becomes final. A suspect does not have to be arrested nor does judicial filing have to have occurred before a suspect registration is made. Registrations may occur before the process of evidence collection has been completed. When a person is declared to be innocent and has been wrongfully suspected of a crime, either by the police during the investigation or by a court, their name is removed from the dataset if this rectification has occurred within three years after the offence has occurred. Court dismissals or the decision not to pursue prosecution will not result in the removal of a suspect from the suspect registrations, except when the reasons for the dismissal was that a person was wrongfully suspected of a crime. The scope of these types of crimes is highly similar to the list of crimes included in the YDS. Many crimes in police statistics are broken down into more detail. For example, the YDS asks in one question about burglary. This offence is disaggregated into several different categories in police records. In addition, the police dataset includes some crimes that are not present in the YDS dataset—such as hostage taking, human trafficking and murder. None of the respondents were suspected of any of these crimes as they rarely occur within this age group. The data on police suspects contains a timeframe of 5 years. The respondents from the 2010 instalment of the YDS were connected to police databases from 2005–2010. The respondents from the 2015 instalment of the YDS were

connected to police databases from 2010–2015. As the police database only goes back as far as 2005, a timeframe of five years was maintained for both the respondents of waves 2010 and 2015. This means that the timeframes included in the YDS does not completely overlap with those of the police suspect registrations.

Social desirability was measured through a scale containing 11 items. Many studies have used sub-sections of the original 33-question MCSDS scale (Beretvas et al. 2002) and variations exist for different target audiences. This version of the MCSDS was translated into the Dutch language and validated by Hermans (1967). It has become a common measure of social desirability for youth in the Netherlands (Hermans 2011). The questions included in the YDS are: (1) I am always honest, (2) I never offend other people, (3) I never boast, (4) I am always friendly to other people, (5) I am always polite to adults, (6) I always admit to my mistakes, (7) I have never intensely disliked somebody, (8) It never bothers me if I am less capable than others, (9) I never try to receive more than others, (10) I never insist on having things go my own way, and (11) I never deliberately said something bad about others.

Analytical Strategies

As mentioned in the introduction, the research question *How can a social desirability measurement be used to increase the validity of data on self-reported criminal behaviour?* is answered by first posing two other questions: *How can social desirability be measured in a way that is comparable for people with different ethnic backgrounds?* And *how does social desirability relate to the underreporting of criminal behaviour?* To address these two preliminary questions, we first examined how to arrive at a measurement of social desirability that relates to the same underlying concepts for people of different ethnic backgrounds. Through the use of a polychoric factor analysis for each separate ethnic group, we compared similarities and differences within the factor structures of this scale. The comparability of the scale was then checked through a measurement invariance test to examine whether systematic differences between groups existed in the scale construct (Van de Vijver and Tanzer 2004; Vandenberg and Lance 2000). We then dropped items with the lowest factor loadings until invariance was established. The second preliminary question was addressed through an analysis of the relationship between the level of social desirability of a respondent and their likelihood to self-report crimes. We separated the respondents into groups based on their social desirability score and displayed for each group the percentage of registered police suspects, the percentage of self-reported crimes, and the percentage of police suspects who also have self-reported a crime. We then compared whether the percentages of these three metrics differed between ethnic groups in order to examine whether these groups differ from each other in the amount of self-reported crime and in the amount of self-reported crime among police suspects.

We addressed the central research question by comparing three different methods that are aimed at minimising bias as a result of the underreporting of criminal activities: (1) by partialling out social desirability in the analytical model, (2) by using listwise deletion, and (3) by using multiple imputation after a procedure we call *social desirability score deletion*.

Partialling out social desirability by using its measure as a control variable is a tool to facilitate comparisons between comparable groups. If respondents are less likely to self-report crimes in an equal measure when they display a high degree of social desirability, then comparing respondents with the same level of social desirability will result

in fair comparisons of criminal behaviour. The exact scope of criminal behaviour may not be found, but the differences in criminal involvement between groups can be compared in this way. This method rests on the assumption that the effect of social desirability on the degree of underreporting is the same across ethnic groups. This method has been recommended by Saunders (1991) and employed, among others, by Weijters et al. (2016), Hermans (1967), Sutton and Farrall (2005) and Leerkes et al. (2019).

The application of listwise deletion is a method to remove respondents who are not deemed to provide accurate information on their own criminal behaviour. This method necessitates some strict assumptions to prevent introducing systematic bias into the dataset. Most importantly, the imposed missingness must be missing completely at random (MCAR). This means that the missing variable does not correlate to another variable. If it is not possible to meet this requirement, the statistical literature concerning the handling of missingness offers a more suitable method to handle unusable responses, which is multiple imputation.

The final method, which we named *Social Desirability based Score Replacement (SDSR)*, is a procedure where the values on self-reported criminal behaviour are removed among respondents with the highest level of social desirability whose scores are not deemed to be comparable across groups in terms of their validity. These values are subsequently replaced through multiple imputation. Multiple imputation is a method that estimates several different values of each missing value, based on auxiliary variables. These are variables that correlate with the missing variable. In our case, the auxiliary variables were: education level of the respondent and their parents; the employment status of the parents; intact family relations; urbanity; income and house value; gender; age; use of alcohol and drugs; hanging around on the street; having delinquent friends; and experiencing social difficulties. Based on this information, the first imputation round was conducted where the values of the missing data were predicted by using stochastic regression analysis. Subsequently, the second round incorporated the new average value and variance to estimate a more exact value for the missing data. In total, between 20 and 30 imputation rounds are recommended (Enders 2010; Johnson and Young 2011). We used 30 imputation rounds in our analysis. Next, all the generated datasets are combined into one dataset where the average of all the imputation scores was calculated. This new dataset could then be used for the main data analysis. Koepfler et al. (2011) and Rios et al. (2017) have used both real data and simulated data to investigate whether listwise deletion or multiple imputation produced more reliable results. Both studies concluded that multiple imputation resulted in the least amount of bias and was therefore the preferred method. Correcting the responses of untrustworthy respondents through multiple imputation has for example been used by Foelber (2017), Kim et al. (2014), Koepfler et al. (2011), and Rios et al. (2017). These researchers used self-imposed missing data based on response times that were quick to the extent that it would not be likely that the respondent had read the whole question. The benefit of multiple imputation is that no respondent needs to be removed from the dataset, allowing the utilisation of a significant part of the information they provided, and preventing a reduction of external validity in the case of the exclusion of a non-random part of the respondents. Second, in contrast to listwise deletion, this method produces no bias, both in the parameters and in the standard errors of the analyses, as long as the missing data are either MCAR or MAR. By including a wide range of auxiliary variables to MI that are related to criminal behaviour, the likelihood of the data being MAR rather than MNAR increases, and thus the appropriateness of MI over traditional missing data methods increases (Foelber 2017).

To compare the impact of these three aforementioned strategies on an analytical model that contains self-reported criminal behaviour, we designed three versions of a regression analysis, each version including a different strategy. We additionally included an analysis variant where social desirability was not taken into consideration. Many studies on self-reported criminal behaviour do not include a social desirability measure or any other strategy to address biases due to underreporting. The inclusion of this fourth condition helps demonstrate to what extent research outcomes are influenced by the inclusion or exclusion of mechanisms to correct underreporting bias. In these analyses, the number of police suspect registrations is used as the dependent variable. This variable is far from normally distributed. Most respondents—5,685 out of 6,218—do not have a suspect registration. The number of registrations of the remaining 533 respondents range from 1 to 14. High amounts of registrations are rare, which makes the variable an over-dispersed count variable. Such skewed distributions in a dependent variable can best be analysed using a *negative binomial regression analysis*. This is a method based on a linear model using maximum likelihood estimation. The scientific notation of this model can be found in Eq. (1).

$$f(x;r,P) = {}^{x-1}C_{r-1} \times P^r \times (1-P)^{x-r} \quad (1)$$

The relation between the independent and dependent variable is expressed with the *incidence rate ratio* (IRR). This measure expresses with $f(x;r,P)$ the probability that in an x number of trials, with r numbers of successes, the probability of success on each trial is P . The IRR therefore does not express absolute values, but it reflects the difference between groups in the predicted number of police registrations. By analysing the relationship between self-reported criminal behaviour and police registrations for different ethnic groups, the varying effects of the three analytic strategies emerge in the outcomes of the analyses.

Because of the large number of zeroes in the dependent variable, we conducted a sensitivity analysis by means of a hurdle model which consists of a two-stage analysis. We first modelled the difference between either having or not having a police registration through a logistic regression. In the second stage, the number of police registrations were modelled among those who have one or more police registration with a negative binomial regression. The outcomes can be found in Attachment 3 of the supplementary materials.

Results

Exploratory Factor Analysis and Measurement Invariance Testing of the Social Desirability Scale

The first step towards answering the research question is by examining how the latent construct of ‘social desirability’ can be measured in a way that is comparable for people with different ethnic backgrounds. In order to address this, we first conducted a polychoric exploratory factor analysis, to accommodate the fact that the items are not continuous but of a binary nature. Table 1 provides an overview of the pattern matrices of the first factor of each ethnic group which were conducted with a Promax rotation method. Separate scree plots for each ethnic group indicated that only one significant dimension was found per group, which was supported by the fact that each group had only one dimension with an eigenvalue exceeding the value of 1.0. These outcomes indicate that each question on the scale relates to the latent variable of social desirability.

Table 1 Structure matrices of the factor analyses per ethnic group of the 11-item social desirability scale (Total $N = 6214$)

Factor structure dimension 1	Native Dutch $N = 2700$	Moroccan origin $N = 648$	Turkish origin $N = 671$	Surinamese origin $N = 682$	Antillean origin $N = 652$	Other origin $N = 861$
1. I am always honest	.6229	.6869	.7225	.6783	.6280	.6495
2. I never offend other people	.7145	.7625	.6556	.7984	.6320	.6894
3. I never boast	.5421	.5726	.6098	.5792	.5969	.5605
4. I am always friendly to other people	.6980	.7319	.6684	.7772	.7337	.7552
5. I am always polite to adults	.5806	.7271	.4121	.6724	.6928	.5843
6. I always admit to my mistakes	.4785	.5557	.4730	.5282	.5146	.5068
7. I have never intensely disliked anyone	.5756	.7433	.7024	.5957	.5710	.5873
8. It never bothers me if I am less capable than others	.3931	.5919	.3735	.4027	.4024	.4865
9. I never try to receive more than others	.5599	.5227	.4481	.5089	.5048	.5445
10. I never insist on having things go my own way	.5212	.6100	.4985	.5240	.4903	.5757
11. I never deliberately said something bad about others	.6604	.7405	.6627	.7502	.6489	.7016

Table 2 Measurement invariance test comparing model fit of social desirability measure where 6 ethnic groups are compared

Measurement invariance test	Items present in the analysis	Chi squared	DF	P	CFI	TLI	RMSEA
Configural	11 items (original scale)	2243.28	264	.000	.812	.766	.085
Configural	10 items (omitting item 8)	1547.37	210	.000	.857	.816	.078
Configural	9 items (omitting item 8 and 5)	106.22	162	.000	.890	.854	.073
Configural	8 items (omitting item 8, 5 and 9)	733.41	120	.000	.915	.881	.070
Configural	7 items (omitting item 8, 5, 9 and 6)	does not concave					
Configural	7 items (omitting item 8, 5, 9 and 10)	492.59	84	.000	.934	.901	.069
Metric	7 items	888.44	114	.000	.875	.862	.081
Configural	6 items (omitting item 8, 5, 9, 10 and 4)	405.25	54	.000	.927	.878	.079
Metric	6 items	641.69	79	.000	.883	.866	.083
Final 6-item model							
Containing item 1, 2, 3, 4, 7 and 11							
Configural	6 items	264.43	54	.000	.960	.933	.061
Metric	6 items	595.35	79	.000	.901	.887	.079
Scalar	6 items	725.92	104	.000	.881	.897	.076
Strict	6 items	1250.98	134	.000	.786	.856	.090

Next, a measurement invariance test was deployed to determine whether the 11-item MCSDS scale is an equally reliable and equally valid measurement of social desirability across ethnic groups. A measure is considered equally reliable across groups if the regression coefficient and the error variances of the observed variables on the latent variables can be constrained to be invariant across groups. In order to determine measurement invariance, a multi-stage process is undertaken. First, a configural invariance test measures the invariance of the model form. Next, a metric invariance test measures the invariance of the item loadings. Thirdly, a scalar invariance test measures the invariance of item intercepts. Finally, a strict invariance test measures the item residual variances. Table 2 displays the outcomes of the measurement invariance tests by presenting the goodness of fit of the models. The Root Mean Square Error of Approximation (RMSEA) reaches an acceptable level below the value of 0.08 (Van de Schoot et al. 2012). The Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) reach an acceptable level when its values exceed 0.9 (ibid.). The first configural invariance test which included the 11-item social desirability scale, indicated that systematic differences remain between ethnic groups. It is recommended in such a case to remove items that disrupt the invariant form of the model (Putnick and Bornstein 2016). We used the previously conducted factor analysis to identify the item with the lowest loading, which was item number 8, and ran the measurement invariance test again without including it. This process was repeated until we reached a model containing 6 items which achieved configural, metric and scalar invariance (see Table 2). The final step, testing invariance of the item residual variances did not reach invariance. However, this is a type of invariance that is rarely attained and the model can be considered sufficiently invariant when scalar invariance is achieved (Putnick and Bornstein 2016; Rosay et al. 2000; Van de Schoot et al. 2012). As a result, it can be concluded that the six-item MCSDS scale produces a sufficiently comparable measurement of the latent construct of

Table 3 Average social desirability scores per ethnic group (N = 6214)

Country of origin	Mean	SD	N
Native Dutch	3.11	1.72	2700
Moroccan origin	4.20	1.72	648
Turkish origin	4.17	1.62	671
Surinamese origin	3.56	1.83	682
Antillean origin	3.45	1.76	652
Other origin	3.51	1.78	861
All respondents	3.48	1.78	6214

social desirability across the ethnic groups present in our sample. The reliability of $\alpha = 0.70$ can be considered to be satisfactory (Nunnally and Bernstein 1994). Barger (2002) has compared different shortened versions of the MCSDS and found that short versions containing as little as five questions are not necessarily worse than longer versions. In some regards, they even function better in terms of model fit (this is the case with the scale used by Hays et al. (1989), for example).

Descriptives of the Six-Item Social Desirability Scale

The six-item social desirability scale that is used from here on out ranges from 0, which is not socially desirable at all, to 6, which is extremely socially desirable. Table 3 shows that the average level of social desirability is 3.48. Native Dutch respondents display the lowest level of social desirability out of the 6 different ethnic groups present in the YDS. Respondents with a Moroccan and Turkish origin have similar levels of social desirability and these two groups display the highest averages. The middle ground is occupied by respondents with a Surinamese and Antillean origin, who also resemble each other in terms of social desirability. An independent sample t-test indicated that there is a significant difference between the degree of social desirability of respondents with a native Dutch background, and respondents with a Moroccan origin $t(3342) = -14.4$, $p = 0.000$, a Turkish origin $t(1075) = -14.89$, $p = 0.000$, a Surinamese origin $t(1005) = -5.79$, $p = 0.000$, an Antillean origin $t(3348) = -4.41$, $p = 0.000$ and other origin $t(1400) = -5.64$, $p = 0.000$.

Associations Between Social Desirability and Criminality

The second step towards answering the research question is to examine how social desirability relates to the underreporting of criminal behaviour. As discussed in the theory section, competing theories exist regarding the meaning of the concept of social desirability. Some studies have found that respondents who score high on this scale are more likely to behave in a socially desirable manner, while others have found that high scores are associated with the underreporting of socially undesirable attitudes and behaviours. To explore the association between social desirability and criminality, an ordered logistic regression analysis is presented in Table 4 with the social desirability scale as dependent variable. The analysis includes the independent variables of ethnic background, age, gender, and ten variables related to self-reported and police-recorded crime. The regression coefficients were also expressed as odds ratios for ease of interpretation. The table confirms that respondents of Moroccan and Turkish descent display the highest level of social desirability.

Table 4 Ordered logistic regression with social desirability as dependent variable (N=6199)

	Coef	SE	Odds ratio	SE
Ethnicity				
Moroccan origin	0.99***	0.08	2.70***	0.22
Turkish origin	0.88***	0.08	2.41***	0.19
Surinamese origin	0.43***	0.08	1.54***	0.12
Antillean origin	0.26**	0.08	1.30**	0.10
Other origin	0.32***	0.07	0.32***	0.10
Native Dutch (ref.)	0.00		1.00	
Age	−0.06**	0.02	0.94**	0.02
Gender (male=0)	0.05	0.05	1.05	0.05
Police registrations	0.12**	0.04	1.13**	0.04
Self-reported offending				
Serious offence	−0.04	0.10	0.96	0.10
Light offence	−0.37***	0.05	0.69***	0.03
Serious drugs	0.14	0.27	1.15	0.31
Light drugs	0.17	0.14	1.18	0.17
Serious property crime	−0.05	0.19	0.95	0.18
Light property crime	−0.24**	0.08	0.78**	0.07
Serious violence	−0.02	0.21	0.98	0.20
Light violence	−0.42***	0.09	0.66***	0.06
Weapon	−0.05	0.14	0.95	0.13
Vandalism	−0.07	0.09	0.93	0.08

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Respondents with a Surinamese, Antillean, and other origin display weaker associations, but the average levels are still significantly higher compared to respondents with a native Dutch background. Younger children report more socially desirable than older youth, but gender is not significantly related to the social desirability scale. The table indicates that there is a small but significant positive relationship between the number of police-suspect registrations and the level of social desirability. Inversely, a significant negative relationship is found between the social desirability scale and self-reporting light offences, particularly property and violent offences.

To further examine the relationship between social desirability and criminality, the respondents were divided into groups based on their social desirability score and migration status. The prevalence in these groups of three different characteristics were subsequently compared: police suspect registrations, self-reported criminal behaviour, and the percentage of police suspects who also self-report criminal behaviour.

The results of this comparison can be found in Fig. 1 (an overview of the exact values presented in this figure with the associated standard deviations and sample sizes is presented in Attachment 1). The figure shows that the percentage of police suspects decreases among respondents with a higher level of social desirability. This decrease however is not significant. The multivariate analysis in Table 4 instead finds a small significant positive association between social desirability and police suspect registrations. Next, Fig. 1 suggests a negative linear association between the level of social desirability and the number of self-reported criminal acts. Table 4 provides more detail regarding this association,

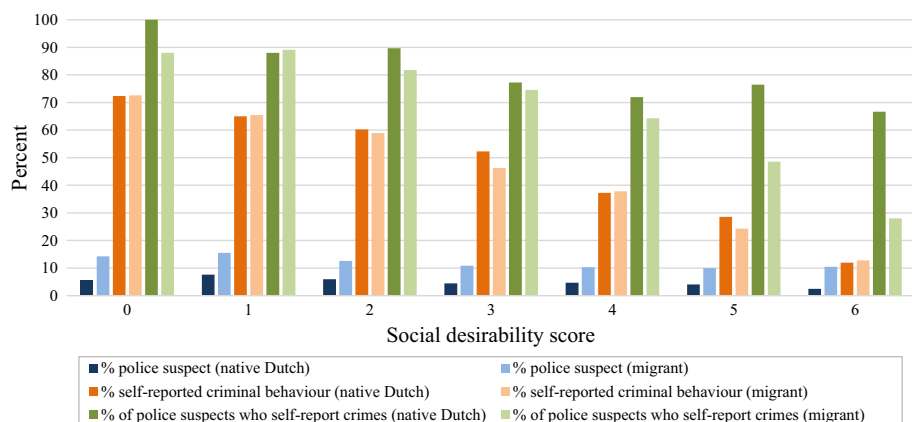


Fig. 1 The percentage of registered suspects and self-reported criminals split by social desirability scores (N=6214)

indicating that it is mainly the result of a significant negative association with light offences, particularly related to property and violent crime. When respondents display a higher degree of social desirability, the likelihood that they will self-report criminal behaviour decreases, but not the likelihood that they become a criminal suspect.

The third variable displayed in Fig. 1 relates to respondents who are both a registered criminal suspect and who self-report at least one criminal act. We cannot assert that each criminal suspect has in fact committed a criminal act, but it would be plausible to assume that the majority of them have. This assumption is supported by the fact that the majority of police suspects who are present in the YDS study also self-report crimes. However, Fig. 1 displays a notable difference in self-reported criminal behaviour among registered police suspects between the values of 0–4 and 5–6. In the range of 0–4, the majority of police suspects self-reported crimes, both those with a native Dutch background ($M=0.84$; $SD=0.37$) and those with a migrant background ($M=0.77$; $SD=0.42$). The difference between these groups is not statistically significant, $t(244.4) = -1.56$, $p=0.121$. Likewise, no significant differences were found when we do not only compare whether the respondent self-reports a crime, but also when we compare the number of self-reported crimes between the groups; native Dutch ($M=6.19$; $SD=7.00$) migrant background ($M=6.30$; $SD=8.88$), $t(361)=0.12$, $p=0.900$. Among police suspects, a negative trend can be discerned between the likelihood to self-report crimes and the level of social desirability, which indicates that for all ethnic groups, the likelihood to self-report a crime when registered as a criminal suspect reduces when the degree of social desirability increases. However, when focusing on the social desirability scores of 5 and 6, registered police suspects with a migration background display a strong decline in the likelihood to self-report a criminal act. This group self-reported criminal acts in only 49% (score 5) and 28% (score 6) of cases. This strong decline in self-reported criminal behaviour among police suspects with a migrant background and a high degree of social desirability is found in each of the major migrant groups living in the Netherlands (see attachment 2 for an overview of these patterns per migrant group). These numbers form a strong contrast with the native Dutch respondents who display a similarly high level of social desirability. This latter group of police suspects self-reported criminal acts in 76.5% (score 5) and 66.7% (score

Table 5 Differences in self-reported offending between the full sample (S1) and a sub-sample (S2) excluding highly socially desirable respondents

Origin	N	Self-reported at least one offence (%)	Average number of light offences	Average number of serious offences
Dutch	N = 2700 (S1)	46.8	1.17 (1.84)	0.13 (0.52)
	N = 2045 (S2)	54.4	1.40 (1.97)	0.16 (0.58)
	Difference	7.6	19.7%	23.1%
Moroccan	N = 648 (S1)	31.1	0.71 (1.51)	0.11 (0.54)
	N = 304 (S2)	47.9	1.22 (1.92)	0.19 (0.73)
	Difference	16.8	71.8%	72.7%
Turkish	N = 671 (S1)	30.7	0.68 (1.45)	0.09 (0.51)
	N = 339 (S2)	44.0	1.03 (1.75)	0.14 (0.67)
	Difference	13.3	51.5%	55.6%
Surinamese	N = 682 (S1)	43.0	1.18 (2.02)	0.14 (0.58)
	N = 434 (S2)	56.5	1.67 (0.19)	0.19 (0.60)
	Difference	13.5	41.5%	35.7%
Antillean	N = 652 (S1)	44.1	1.06 (1.83)	0.19 (0.81)
	N = 441 (S2)	52.5	1.35 (2.02)	0.24 (0.95)
	Difference	8.4	27.4%	26.3%
Other	N = 861 (S1)	40.8	1.01 (1.85)	0.18 (0.79)
	N = 570 (S2)	52.5	1.35 (2.09)	0.25 (0.93)
	Difference	11.7	33.7%	38.9%

6) of cases. According to an independent sample *t* test, this difference between native Dutch ($M = 0.74$; $SD = 0.45$) and respondents with a migrant background ($M = 0.38$; $SD = 0.49$) is highly significant: $t(30.8) = -3.53$, $p = 0.001$. This finding is an indication that a comparison of self-reported criminal behaviour between respondents belonging to an ethnic minority group and the ethnic majority group is not valid within the highest range of the social desirability scale. Therefore, a dataset which only includes respondents with a social desirability score of 4 and below would be more suitable to make comparisons between ethnic groups.

Table 5 displays the average levels of self-reported criminal behaviour per ethnic group. A comparison was made between the means of all respondents per group and the means of the subsample of the group where respondents with a social desirability score of 5 and 6 had been excluded. The table includes information on the prevalence of self-reported crime, which refers to the percentage of respondents who reported to have ever committed at least one crime. It also includes the average self-reported frequencies for light and serious crimes. The differences in the prevalence and the frequency of self-reported crime between the two samples are smallest among respondents with a native Dutch origin, closely followed by respondents with an Antillean origin. Respondents with a Surinamese and other background take the middle ground, and the largest changes can be observed among respondents with a Moroccan or Turkish origin. Concurrently, these latter two groups also display the highest levels of social desirability. The amount of light and serious offences increased most strongly for the respondents with a Moroccan origin when comparing the subsample S2 with the full sample S1.

The averages in the subsample increased over 70% compared to the full sample. These changes form a stark contrast with the native Dutch group, for whom the number of self-reported light and serious offences only increased around 20%. As a result, the level of self-reported offending of the different ethnic groups is much more similar in the subsample as compared to the full sample.

Comparing Different Correction-Techniques to Minimise Social Desirability bias in Regression

Having arrived at a measurement of social desirability that holds a comparable meaning for people with different ethnic backgrounds, and having examined the relationship between social desirability and the underreporting of criminal behaviour, we can progress towards answering the central research question: *How can a social desirability measurement be used to increase the validity of data on self-reported criminal behaviour?* In this section, three different methods aimed at minimising underreporting bias related to social desirability are compared. This assessment takes place through the comparison of four variations of a regression analysis. In the first variant, underreporting bias is not taken into account. The second variant controls for social desirability. The third variant excludes underreporters through listwise deletion. The fourth and final variant is preceded by the removal of the self-reported criminal behaviour scores among respondents who scored a 5 or higher on the social desirability scale. Next, multiple imputation techniques were used aiming to correct underreporting bias by estimating accurate scores of the self-reported criminal behaviour based on the patterns among the respondents with low to average levels of social desirability.

Table 6 displays the four analysis variants in which the number of police registrations of criminal suspects is predicted using negative binomial regression. Each analysis contains two models. The first model contains the independent variables of ethnic background, age, and gender. The second model contains additional variables on self-reported criminal behaviour. The outcomes of these analyses are presented through the Incidence Rate Ratio (IRR), which expresses the relative likelihood that the dependent variable occurs. An IRR of 6.56 for respondents with a Moroccan origin, for example, indicates that respondents from this group on average have 6.56 times as many police registrations compared to respondents from the reference group, which consists of youth with a native Dutch background.

In the first variant of this analysis, social desirability is not taken into consideration. Its outcomes indicate very large differences in the predicted number of police suspect registrations between native Dutch respondents and the ethnic minority groups. Respondents with an Antillean background have the most registrations—7.27 times more than native Dutch respondents. The smallest number of police registrations among the different migrant groups can be found in the residual group, which consists of respondents with a migrant background from countries other than Morocco, Turkey, Surinam, or the Dutch Antilles. Their average number of police registrations is still 3.07 times as high compared to the native Dutch group. When self-reported criminal behaviour is added in model 2, the difference in police registrations between the native Dutch respondents and the respondents with a Moroccan and Turkish background shows a small but non-significant increase. A small non-significant decrease is found among the other three groups. These outcomes would suggest that variations in self-reported offending cannot explain any part of the overrepresentation of ethnic minorities in police suspect registrations.

Table 6 Negative Binomial regression models

	Variant 1: no corrections						Variant 2: SD as control						Variant 3: listwise deletion						Variant 4: SDR					
	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6			Model 7			Model 8		
	IRR	SE		IRR	SE		IRR	SE		IRR	SE		IRR	SE		IRR	SE		IRR	SE		IRR	SE	
<i>Migration background</i>																								
Moroccan origin	6.56***	1.08		7.20***	1.16		7.02***	1.18		6.54***	1.06		6.24***	1.33		5.66***	1.14		6.56***	1.08		5.91***	0.97	
Turkish origin	4.63***	0.78		5.05***	0.85		5.01***	0.87		4.61***	0.79		4.24***	0.92		3.91***	0.84		4.63***	0.78		4.42***	0.76	
Surinamese origin	4.61***	0.78		4.60***	0.77		4.79***	0.82		4.38***	0.73		4.44***	0.89		3.69***	0.73		4.61***	0.78		4.07***	0.69	
Antillean origin	7.27***	1.19		7.09***	1.13		7.49***	1.24		6.84***	1.08		6.22***	1.18		5.81***	1.05		7.27***	1.19		6.32***	1.01	
Other origin	3.07***	0.50		2.84***	0.46		3.20***	0.52		2.71***	0.44		2.84***	0.53		2.23***	0.42		3.07***	0.50		2.55***	0.42	
Native Dutch	1			1			1			1			1			1			1			1		
Age	2.43***	0.14		1.99***	0.11		2.43***	0.14		1.97***	0.11		2.14***	0.15		1.69***	0.11		2.43***	0.14		1.91***	0.11	
Gender (male = 0)	0.29***	0.03		0.35***	0.04		0.30***	0.03		0.35***	0.04		0.30***	0.04		0.40***	0.05		0.29***	0.03		0.38***	0.04	
Self-reported criminal behaviour																								
No. Serious offences				1.14	0.18					1.15	0.17					1.15	0.18					1.19	0.19	
No. Light offences				1.16	0.10					1.21*	0.10					1.21	0.11					1.04	0.10	
Drugs offence serious				1.03	0.37					1.06	0.38					1.17	0.36					0.92	0.36	
Drugs offence light				1.29	0.28					1.24	0.27					1.33	0.30					1.23	0.29	
Property crime serious				2.18**	0.62					2.18**	0.62					2.58**	0.69					2.26**	0.70	
Property crime light				1.09	0.17					1.09	0.17					0.84	0.17					1.09	0.19	
Violence serious				1.31	0.37					1.35	0.38					1.39	0.40					1.28	0.36	
Violence light				1.25	0.21					1.29	0.21					1.41*	0.26					1.47*	0.29	
Weapon offence				1.08	0.24					1.06	0.23					0.99	0.24					1.14	0.26	
Vandalism				1.01	0.15					1.01	0.15					0.95	0.17					1.11	0.18	
Social desirability							0.94	0.03		1.10***	0.03													
Constant	0.06***	0.01		0.03***	0.00		0.07***	0.01		0.02***	0.00		0.07***	0.01		0.03***	0.00		0.06***	0.01		0.00***	0.00	
Model characteristics																								
N	6218			6214			6203			6199			4136			4133			6218			6218		

Table 6 (continued)

	Variant 1: no corrections				Variant 2: SD as control				Variant 3: listwise deletion				Variant 4: SDSR			
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
	IRR	SE	IRR	SE	IRR	SE	IRR	SE	IRR	SE	IRR	SE	IRR	SE	IRR	SE
Pseudo R squared	0.106		0.150		0.108		0.153		0.091		0.150					
Log likelihood	−2164		−2056		−2157		−2046		−1483		−1386					
Alpha	4.58	0.41	3.14	0.31	4.57	0.30	3.04	0.30	4.89	0.51	2.85	0.35	4.58	0.41	2.93	0.31

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

The second variant of the analysis partialles out the effect of social desirability through its inclusion as a control variable. Initial levels of overrepresentation as displayed in model 3 are significantly higher than those in model 1.³ The difference between models 3 and 4 are more pronounced in this second variant compared to models 1 and 2. When self-reported criminal behaviour is added to the analysis, the amount of overrepresentation has decreased the most for respondents in the residual group (15.3%). The second largest reduction is found among respondents with an Antillean background (8.7%). Smaller reductions appear among respondents of Surinamese (8.6%), Turkish (8.0%) and Moroccan (6.8%) origin. Postestimation tests however found these differences to be statistically non-significant.

In the third variant of the analysis, all respondents who scored either a 5 or 6 on the social desirability scale were omitted through listwise deletion. This method did not result in a significant change of initial levels of overrepresentation as seen in the second variant. When comparing models 2 and 6, overrepresentation of respondents with an Antillean origin significantly lowered with 22.0% ($X^2(1)=3.73$, $p=0.05$) and for respondents with an other origin it significantly lowered with 21.5% ($X^2(1)=3.99$, $p=0.04$). The other changes, although substantial, are not statistically significant; 27.23% for respondents with a Moroccan origin; 29.2% for respondents with a Turkish origin and 24.7% for respondents with a Surinamese origin. The analytical variant that is chosen, has therefore a significant effect on the amount of overrepresentation that is found. It also has a small but significant effect on the amount of overrepresentation that can be explained by self-reported criminal behaviour. Out of the three methods, listwise deletion seems to achieve the most valid estimates due to the methodological reasons as laid out in this paper. However, this variant requires the removal of all the respondents with high social desirability scores and will therefore introduce significant bias in the dataset. In addition, a large percentage of police suspects with an ethnic minority status (33.4%) is lost when respondents with a high degree of social desirability are removed from the dataset. It would be more desirable to keep these respondents while correcting for underreporting bias. Through multiple imputation, this aim can be achieved.

The fourth and final variant utilises our method of *Social Desirability based Score Replacement* (SDSR) to estimate valid values of criminal behaviour among respondents who scored either a 5 or 6 on the social desirability scale. The initial degree of overrepresentation displayed in model 7 is the same as in model 1, because the variables of ethnicity, gender, and age are not subject to imputation. In this variant, self-reported criminal behaviour accounts for 9.9% (Moroccan origin), 4.3% (Turkish origin), 11.7% (Surinamese origin), 13.1% (Antillean origin) and 16.9% (other origin) of the overrepresentation of these ethnic minority groups. When self-reported criminal behaviour is added in this variant, the difference in police registrations decreases between the native respondents and the respondents with a migrant background, indicating that a small part of the overrepresentation of ethnic minority groups (4–17%) can be explained by differences in (self-reported) criminal behaviour.

The values of the variables related to ethnicity vary dependent on the analytic variant that was used. Significant differences were found between each ethnic group in model 2 and 6, but a comparison between model 2 and 4 only resulted in significant differences among

³ A significant increase is found among respondents with a Turkish background $X^2(1)=4.77$, $p=0.03$, and an other background $X^2(1)=4.26$, $p=0.04$. Marginally significant changes are found for respondents with a Moroccan background $X^2(1)=3.51$, $p=0.06$ and a Surinamese background $X^2(1)=3.37$, $p=0.07$.

the groups with the largest sample sizes, namely the other group and the Antillean group. We therefore considered that the sample size might have influenced the significance levels of coefficient changes. For this reason, we ran the models after combining several ethnic minority group samples. Patterns of police contact were similar for youth with a Moroccan and Antillean origin. Likewise, youth with a Turkish and Surinamese origin resembled each other in terms of criminal overrepresentation. When these four groups were merged into two, each analytical model which contains the measurement of self-reported criminal behaviour was significantly different as compared to the first variant where no corrections had been applied.

Discussion

Self-reported criminal behaviour is a data source with the potential to expand our understanding of patterns of criminal behaviour, because it can provide insight into criminal behaviour that takes place both inside and outside the view of the police. These data can additionally facilitate insight into the question whether some groups have a disproportionate likelihood of getting into contact with the police after committing a crime. Police statistics are continued to be used as a proxy for criminal behaviour, even though a biased selection process takes place between criminal *behaviour* and criminal *registration*. Past studies on self-reported criminal behaviour have not been able to reproduce the substantial differences between ethnic groups as found in police statistics. It is important to ensure that these differences between police statistics and self-reported criminal behaviour are not the result of validity issues in self-reported data, as past studies found that ethnic minorities living in Europe and North America have a greater propensity to underreport criminal behaviour (Bersani and Piquero 2017; Hindelang et al. 1979; Junger 1989; Kirk 2006; Lab and Allen 1984). Methodological innovations regarding self-reported criminal behaviour are therefore warranted to minimise systematic differences in underreporting bias and improve intergroup comparability. To improve the accuracy of data on criminal behaviour derived from self-reports, we posed the following research question: *How can a social desirability measurement be used to increase the validity of data on self-reported criminal behaviour?* However, social desirability scales have not been designed for people of different ethnic origins and its outcomes may therefore not be comparable. Furthermore, the exact relationship between social desirability and the underreporting of criminal behaviour is not clear. Before we could address our central research question, two other questions therefore needed to be addressed: *How can social desirability be measured in a way that is comparable for people with different ethnic backgrounds?* And: *How does social desirability relate to the underreporting of criminal behaviour?*

A measurement invariance test revealed that the commonly used 11-item social desirability scale did not reap comparable results for the six different ethnic groups in our sample. This suggests that the meaning of social desirability as originally envisioned by Marlowe and Crowne is in fact not a universal one. We have therefore removed the lowest scoring factors determined by a polychoric factor analysis and arrived at a six-item scale with a sufficient level of invariance. This answered our first question on how to measure social desirability in a way that is comparable for people with different ethnic backgrounds.

Subsequently, the relation between social desirability and the underreporting of criminal behaviour was examined. In a multivariate analysis with controls for gender, age, and ethnic background, a small but significant positive association was found between social

desirability and the amount of police registrations. A much stronger negative relationship was found between social desirability and self-reported crime. This outcome suggests that social desirability is more likely to be associated with the underreporting of criminal behaviour instead of with a lower likelihood to commit crimes.

Next, patterns of self-reported offending were compared among registered crime suspects with different social desirability scores. Substantial and significant differences were found between suspects with and without a migrant background in the associations between social desirability and self-reported crime. The majority of the native-Dutch crime suspects also self-reported at least one crime, which indicates that the validity of self-reported criminal behaviour in this group is high. Similarly, criminal suspects with a migrant background (including those with a western and a non-western origin) also largely reported at least one crime when they had a low-to-medium level of social desirability. After the listwise deletion of respondents with the highest social desirability scores, an independent sample t-test showed no significant differences between ethnic groups in terms of self-reported criminal behaviour. In contrast, registered crime suspects with both an ethnic minority status and a high social desirability score were not likely to self-report any crime, which points towards a propensity to underreport criminal behaviour. A second t-test comparing respondents with a high level of social desirability revealed significant differences between the ethnic minority and ethnic majority group in the self-reporting of criminal behaviour. This finding suggests that the negative association between social desirability and self-reported criminal behaviour among ethnic minority respondents can be interpreted as an indicator that social desirability affects the willingness to *disclose* undesirable behaviour instead of it being an indicator of the *display* of undesirable behaviour. An alternative explanation could be that police suspects with both an ethnic minority background and a high degree of social desirability have a higher likelihood of being wrongfully suspected of a crime. This explanation however does not find strong support, particularly because suspects who eventually have been found to be innocent were no longer marked as a registered suspect in the dataset.

The negative association between self-reported criminal behaviour and social desirability may warrant a different explanation for native Dutch youth. In this group, the decreased likelihood to self-report crimes was accompanied by a decreased likelihood of being suspected of a crime by the police. In addition, criminal suspects largely self-reported crimes, even among respondents with very high levels of social desirability. This suggests that social desirability among native Dutch respondents may be related to the display of criminal behaviour, instead of the willingness to disclose criminal behaviour.

These findings indicate the existence of systematic differences between ethnic groups in the way that social desirability relates to underreporting. It seems that underreporting mostly occurs among respondents with both a migrant background and a high level of social desirability. We found these patterns among respondents with a Moroccan, Turkish, Surinamese and Antillean origin, but also among the residual group, which consists of people with both a European and non-European migration background. Underreporting seemed to be particularly prevalent among respondents who originate from Morocco and Turkey. A possible explanation for this finding might be that respondents are less likely to report negative behaviour when they originate from a place where a so-called “honour and shame culture” exists. For a person socialised in such a culture, the infliction of shame upon themselves—and by association also on the group they belong to—has much more serious implications compared to people who have been socialised in more individualistic “guilt” cultures (Hermans 1999; Peristiany 1965). This makes the concealment of unfavourable traits and behaviours more socially acceptable and expected among people who live in a shame culture, compared to people living in a guilt culture (Klooster et al. 1999). Another reason for the difference between native

Dutch respondents and respondents with a non-western migrant background in the propensity to underreport criminal behaviour may lie in the different social position that these two groups occupy in society. It is possible that people with a more uncertain social status feel less confident to admit to undesirable behaviour (Lalwani et al. 2006). We conclude that respondents with and without a migrant background who display a high level of social desirability (a score of 5 or 6) are not deemed to be comparable in terms of the validity of their self-reported criminal behaviour.

The central research question on how to use the social desirability measure to increase the validity of self-reported criminal behaviour was addressed by comparing four variations of a negative binomial regression analysis. Its outcomes indicate that paying no attention to social desirability leads to an underestimation of criminal behaviour among respondents with an ethnic minority status, especially Moroccan and Turkish respondents. These groups are more likely to underreport criminal behaviour when they score high on the social desirability scale. Controlling for social desirability therefore does not resolve this problem, since the level of underreporting among socially desirable respondents who display high levels of social desirability does not occur to a comparable degree. Using this method will therefore lead to an overestimation of the differences between ethnic groups in terms of police suspect registrations. Simultaneously, it leads to an underestimation of the effect that self-reported criminal behaviour has on the likelihood to be regarded as a criminal suspect. Listwise deletion removes the incongruence in validity of self-reported criminal behaviour among respondents with a high level of social desirability. However, this method introduces a strong degree of bias in the data. Not only does it result in the loss of a large portion of respondents, but it also leads to a significant underestimation of the initial degree of overrepresentation of ethnic minority groups. We therefore conclude that the method of *Social Desirability Related Score Replacement* results in the most valid outcomes of self-reported criminal behaviour since it targets validity issues exclusively among respondents with systematically different levels of underreporting without removing them from the analysis. Since the patterns of missingness from the self-imposed missing data were not completely random, this method will unavoidably also introduce a certain degree of bias. In addition, a potential risk lies in the chance that some people with a higher degree of social desirability might report less crimes because they are less likely to commit criminal acts. If this is the case, replicating the patterns of self-reported criminal behaviour among respondents with a low to medium degree of social desirability onto respondents with a high degree of social desirability would lead to an overestimation of criminal behaviour. Arriving at an accurate measure of criminal behaviour therefore remains a challenge. However, the application of the novel method that we call SDSR, is not aimed at revealing exact levels of criminal behaviour. Instead, it facilitates valid intergroup comparisons which is useful for theory-development, especially in relation to ethnic inequalities in judicial processing.

Limitations and Recommendations for Further Research

This paper is aimed to garner renewed interest in the methodological improvement of the measurement of self-reported criminal behaviour. Focussing on the topic of social desirability, practical recommendations are provided to prevent bias as a result of the underreporting of criminal behaviour. The original MDSDS consists of 33 questions, which may deter researchers from including it in their surveys (MacDonald et al. 2017). Instead, we recommend using a shorter scale such as the 6-item scale presented in this paper which is

demonstrated to be invariant across multiple ethnic groups. Researchers have been using different approaches to include social desirability measurements in their models and the outcomes of these models are strongly impacted by these choices. We therefore recommend that untrustworthy responses or otherwise unusable values are better resolved through our method of *Social Desirability based Score Replacement* compared to the methods of list-wise deletion, partialling out social desirability, or by ignoring underreporting bias altogether. The method of SDSR facilitates intergroup comparison, introduces less bias into the dataset and preserves valuable information that would otherwise be omitted.

However, it ought to be emphasised that the validity of self-reported criminal behaviour does not solely hinge on the underreporting of criminal behaviour, and can be affected by many other factors, such as non-response (Weijters et al. 2016), overreporting (Krohn et al. 2013), forgetfulness (Short et al. 2009) or differences in self-control (MacDonald et al. 2017; Piquero et al. 2000). For these reasons, the inclusion of a social desirability measure cannot solve all validity issues of self-reported criminal behaviour. Instead, it can be used to address one particular cause of bias and may therefore contribute towards the accurate measurement of self-reported criminal behaviour among different ethnic groups. More generally, the inclusion of a social desirability measure and its suitable use to minimise bias may improve the measurement of self-reported attitudes and behaviours in other fields as well that relate to socially undesirable or stigmatized behaviour, for instance, sexual behaviour, bullying or harsh parenting. It may furthermore facilitate comparisons between other groups that likewise show systematic variations in the amount of social desirability, such as people differing in terms of age or social class.

Another limitation regarding the accuracy of the data on self-reported criminal behaviour relates to the respondents' understanding of the legal demarcation of criminal offences. Our strongest concern in this regard relates to the offence 'assault without injury'. This offence was most often self-reported by the respondents. However, it is not fully clear what type of incidences respondents have grouped under this category and it is possible that innocent children's behaviour was interpreted as assault without injury. We therefore recommend the addition of a short explanation to accompany certain offences, to help respondents understand which behaviours do and which do not fit in the criminal categories that are included in the questionnaire.

We finally recommend critical examinations of existing social desirability measures and the development of adjusted or new scales which can facilitate inter-group comparisons. Our factor analyses of different ethnic groups show important differences in the internal structure of the commonly used 11-item social desirability scale, which complicates the comparison between these groups. As with many psychological scales, the original Marlowe-Crowne social desirability scale was developed using so-called WEIRD data, meaning that the scale was designed by and for a western, educated, industrialised, rich, democratic group of people. More specifically, this scale—which is still the most common measure of social desirability—was validated and tested by a sample of psychology students from Ohio State University in 1960 (Crowne and Marlowe 1960). The assumption that these scales have a comparable meaning for people across different cultures is not supported in this paper. We therefore recommend that research into sensitive topics such as criminal behaviour and other forms of deviance can be improved by the inclusion of a proper social desirability scale, as the inclusion of such a scale can serve to reduce systematic differences in underreporting bias. Self-reported data are an important source of information in the social sciences, but serious attention should be paid to the validity of this type of data.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10940-023-09567-y>.

Funding Partial financial support was received from Police and Science (grant no. Z801110 / Project PW/OC/2019/20).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barger SD (2002) The Marlowe–Crowne affair: short forms, psychometric structure, and social desirability. *J Pers Assess* 79(1):286–305
- Beretvas SN, Meyers JL, Leite WL (2002) A reliability generalization study of the Marlowe–Crowne social desirability scale. *Educ Psychol Measur* 62(1):570–589
- Bersani BE, Piquero AR (2017) Examining systematic crime reporting bias across three immigrant generations: prevalence, trends, and divergence in self-reported and official reported arrests. *J Quant Criminol* 33(1):835–857
- Boon A, van Dorp M, de Boer S (2019) Disproportionate minority contact in the Dutch juvenile justice system. *Journal of Ethnicity in Criminal Justice* 17(1):42–56
- Buil-Gil D, Medina J, Shlomo N (2021) Measuring the dark figure of crime in geographic areas. Small area estimation from the crime survey for England and Wales. *British J Criminol* 61(2):364–388
- Centraal Bureau voor de Statistiek (2010) *Integrale Veiligheidsmonitor 2009*. Centraal Bureau voor de Statistiek, Den Haag
- Crowne DP, Marlowe D (1960) A new scale of social desirability independent of psychopathology. *J Consult Psychol* 24(4):349–354
- Elffers H, van der Kemp JJ (2016) *Cahiers politiestudies: meten is weten in als je nou politiecijfers combineert met slachtofferenquêtes, dan ben je nog nergens wat nu?* essay. MAKLU Uitgevers 41:43–55
- Enders CK (2010) *Applied missing data analysis*. Guilford Press, New York
- Engbersen G, van der Leun JP, de Boom J (2007) The fragmentation of migration and crime. *Crime Justice* 35(1):389–452
- Farrington DP, Loeber R, Stouthamer-Loeber M, Van Kammen WB, Schmidt L (1996) Self-reported delinquency and a combined delinquency seriousness scale based on boys, mothers, and teachers: concurrent and predictive validity for African-Americans and Caucasians. *Criminology* 34:493–517
- Fernandes MF, Randall DM (1992) The nature of social desirability response effects in ethics research. *Bus Ethics Q* 2(2):183–205
- Foelber KJ (2017) Using multiple imputation to mitigate the effects of low examinee motivation on estimates of student learning. *Dissertations* 148, <https://commons.lib.jmu.edu/diss201019/148>.
- Goff PA, Jackson MC, Di Leone BAL, Culotta CM, DiTomasso NA (2014) The essence of innocence: consequences of dehumanizing black children. *J Pers Soc Psychol* 106(4):526–545
- Grisso T, Schwartz RB (eds) (2000) *Youth on trial; a developmental perspective on Juvenile justice*. University of Chicago Press, Chicago
- Hanson RK, Bussière MT (1998) Predicting relapse: a meta-analysis of sexual offender recidivism studies. *J Consult Clin Psychol* 66:348–362
- Hays RD, Hayashi T, Stewart AL (1989) A five-item measure of socially desirable response set. *Educ Psychol Measur* 49:629–636
- He J (2015) *The general response style from a cross-cultural perspective* (Dissertation). Tilburg University, Tilburg
- Hermans P (1999) The expression of guilt by Moroccan adolescents: ethnocentric interpretations by western teachers and social workers. *Int J Educ Res* 31:303–316
- Hermans HJM (1967) *Motivatie en prestatie* (dissertation). Swets and Zeitlinger.

- Hermans HJM (2011) PMT-K-2 prestatie motivatie test voor kinderen (2de ed). Pearson etc.
- Hindelang MJ, Hirschi T, Weis JG (1979) Correlates of delinquency: the illusion of discrepancy between self-report and official measures. *Am Sociol Rev* 44(6):995–1014
- Jennissen RPW, Blom M (2007) Allochtone en autochtone verdachten van verschillende delicttypen nader bekeken. Den Haag, WODC/CBS.
- Johnson TP, van de Vijver FJR (2003) Social desirability in cross-cultural research. In: Harkness JA, van de Vijver FJR, Mohler PP (eds) Cross-cultural survey methods. Wiley, New Jersey, pp 195–204
- Johnson DR, Young R (2011) Toward best practices in analyzing datasets with missing data: comparisons and recommendations. *J Marriage Fam* 73:926–945
- Jolliffe D, Farrington DP, David Hawkins J, Catalano RF, Hill KG, Kosterman R (2003) Predictive, concurrent, prospective and retrospective validity of self-reported delinquency. *Crim Behav Mental Health* 13:179–197
- Junger M (1989) Discrepancies between police and self-report data for Dutch racial minorities. *Br J Criminol* 29:273–284
- Junger M (1990) Delinquency and ethnicity: an investigation on social factors relating to delinquency among Moroccan, Turkish Surinamese and Dutch boys. Brill Archive, Dutch
- Junger-Tas J, Marshall IH, Enzmann D, Killias MW, Steketee M, Gruszczynska B (2010) Juvenile delinquency in Europe and beyond: results of the second international self-report delinquency study. Springer, New York
- Kashy DA, DePaulo BM (1996) Who lies? *J Pers Soc Psychol* 70:1037–1051
- Kim HJ, Reiter JP, Wang Q, Cox LH, Karr AF (2014) Multiple imputation of missing or faulty values under linear constraints. *J Bus Econ Stat* 32(3):375–386
- Kirk DS (2006) Examining the divergence across self-report and official data sources on inferences about the adolescent life-course of crime. *J Quant Criminol* 22:107–129
- Kivivuori J (2011) Discovery of hidden crime: self-report delinquency surveys in criminal policy context. Oxford University Press, Oxford
- Klooster EM, Hoek AJE, Hoff van CAT (1999). Allochtonen en strafbeleving: een onderzoek naar de strafbeleving van antilliaanse, surinaamse, marokkaanse en turkse jongens, directie preventie, jeugd en sanctiebeleid (PJS), Den Haag.
- Koepfler J, Jurich D, DeMars C (2011) Applying modern missing data techniques to the treatment of rapid responses. In: paper presented at the annual meeting of the 121 National Council on Measurement in Education, New Orleans.
- Krohn MD, Lizotte AJ, Phillips MD, Thornberry TP, Bell KA (2013) Explaining systematic bias in self-reported measures: factors that affect the under- and over-reporting of self-reported arrests. *Justice Q* 30:501–528
- Krumpal I (2013) Determinants of social desirability bias in sensitive surveys: a literature review. *Qual Quant* 47:2025–2047
- Kurlychek MC, Johnson BD (2019) Cumulative disadvantage in the American criminal justice system. *Annu Rev Criminol* 2(1):291–319
- Laan van der AM, Blom M (2006) Jeugdcriminaliteit: risico's en bescherming—bevindingen uit de WODC Monitor Zelfgerapporteerde Jeugdcriminaliteit 2005. Den Haag: Boom Juridische uitgevers. Onderzoek en beleid 245
- Lab SP, Allen RB (1984) Self-report and official measures: a further examination of the validity issue. *J Crim Justice* 12:445–455
- Lalwani AK, Shavitt S, Johnson T (2006) What is the relation between cultural orientation and socially desirable responding? *J Pers Soc Psychol* 90:165–178
- Leerkes A, Martinez R, Groeneveld P (2019) Minority paradoxes: ethnic differences in self-reported offending and official crime statistics. *Br J Criminol* 59(1):166–187
- Lew ED (1997) Criminaliteit en etnische minderheden; Een criminologische analyse, Ministerie van Justitie, WODC, K-reeks (51), Den Haag
- MacDonald JM, Morral A, Piquero AR (2017) Socially desirable response bias in criminology: an example of its effect in testing the effects of self-control. In: MacDonald JM (ed) Measuring crime and criminality. Taylor and Francis, London, pp 21–36
- Maxfield MG, Babbie E (1998) Research methods for criminal justice and criminology, 2nd edn. Wadsworth, Belmont
- Maxfield MG, Weller BL, Widom CS (2000) Comparing self-reports and official records of arrests. *J Quant Criminol* 16:87–110
- Mills JF, Kroner DG (2005) An investigation into the relationship between socially desirable responding and offender selfreport. *Psychol Serv* 2:70–80
- Nunnally JC, Bernstein IH (1994) Psychometric theory, 3rd edn. McGraw-Hill, New York

- Paulhus DL (1981) Control of social desirability in personality inventories: principal-factor deletion. *J Res Pers* 15:383–388
- Paulhus DL (1984) Two-component models of socially desirable responding. *J Pers Soc Psychol* 46:598–609
- Peristiany JG (1965) Honour and shame: the values of Mediterranean society. Weidenfeld and Nicolson, London
- Piquero AR, Brame R (2008) Assessing the race-/ethnicity-crime relationship in a sample of serious adolescent delinquents. *Crime Delinq* 54:390–422
- Piquero AR, MacIntosh R, Hickman M (2000) Does self-control affect survey response? Applying exploratory, confirmatory, and item response theory analysis to Grasmick et al.'s self-control scale. *Criminol* 38(3):897–930
- Putnick D, Bornstein M (2016) Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev Rev* 41:71–90
- Randall DM, Paul Huo Y, Pawelk P (1993) Social desirability bias in cross-cultural ethics research. *The Int J Organ Anal* 1(2):185–202
- Rios JA, Guo H, Mao L, Liu OL (2017) Evaluating the impact of careless responding on aggregated-scores: to filter unmotivated examinees or not? *Int J Test* 17:74–104
- Rosay AB, Gottfredson DC, Armstrong TA et al (2000) Invariance of measures of prevention program effectiveness: a replication. *J Quant Criminol* 16:341–367
- Ross CE, Mirowsky J (1984) Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *J Health Soc Behav* 25(2):189–197
- Saunders D (1991) Procedures for adjusting self-reports of violence for social desirability bias. *J Interpers Violence* 6:336–344
- Sellin T (1931) The basis of a crime index. *J Crim Law Crim* 22:335–356
- Short ME, Goetzel RZ, Pei X, Tabrizi MJ, Ozminkowski RJ, Gibson TB, Dejoy DM, Wilson MG (2009) How accurate are self-reports? analysis of self-reported health care utilization and absence when compared with administrative data. *J Occup Environ Med* 51(7):786–796
- Sugarman DB, Hotaling GT (1997) Intimate violence and social desirability. *J Interpers Violence* 12:275–290
- Sutton R, Farrall S (2005) Gender, socially desirable responding and the fear of crime. are women really more anxious about crime. *Br J Criminol* 45:212–224
- Tan L, Grace RC (2008) Social desirability and sexual offenders: a review. *Sex Abuse J Res Treat* 20:61–87
- Thornberry T, Krohn M (2000) The self report method for measuring delinquency and crime. *Meas Anal Crime Justice* 4:33–83
- Tracy PE (1987) Race and class differences in official and self-reported delinquency. In: Wolfgang ME, Thornberry TP, Figlio RM (eds) *From boy to man, from delinquency to crime*. University of Chicago Press, Chicago, pp 87–121
- Twenge JM, Im C (2007) Changes in the need for social approval: 1958–2001. *J Res Pers* 41:171–189
- Uziel L (2010) Rethinking social desirability scales: from impression management to interpersonally oriented self-control. *Perspect Psychol Sci* 5:243–262
- van de Schoot R, Lugtig PJ, Hox JJ (2012) Methodology, and statistics for the behavioural and social sciences, and Afd, methoden en statistieken. A checklist for testing measurement invariance. *Eur J Dev Psychol* 9(4):486–492
- van der Leun, J, Kromhout M, Easton M, Weerman F (2010) Criminaliteit, migratie en etniciteit: nieuwe richtingen binnen een complex en beladen onderzoeksterrein. *Tijdschr Criminol* 52(2):107–107
- Vandenberg RJ, Lance CE (2000) A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ Res Methods* 3(1):4–69
- Verbruggen J, Blokland AAJ, Robinson AL, Maxwell CD (2021) General offending and intimate partner violence perpetration in young adulthood: a Dutch longitudinal study. *Int J Offender Ther Comp Criminol*. <https://doi.org/10.1177/0306624X211022657>
- Vijver van deTanzer FNK (2004) Bias and equivalence in cross-cultural assessment: an overview. *Eur Rev Appl Psychol* 54(2):119–135
- Weijters G, Laan AM van der, Kessels JR (2016) De overeenstemming tussen zelfgerapporteerde jeugd-criminaliteit en bij de politie bekende jeugdige verdachten. *WODC, Cahier* 3
- Xie M, Lauritsen J (2012) Racial context and crime reporting: a test of black's stratification hypothesis. *J Quant Criminol* 28:265–293