School vs. mail surveys: disentangling selection and measurement effects in self-reported juvenile delinquency

Diederik Cops, Arne De Boeck & Stefaan Pleysier

Leuven Institute of Criminology

Introduction

From the origin of criminology as a scientific discipline, the measurement of crime and criminal behaviour has been and still is a pivotal topic, but at the same time a source of continuous frustration. This is largely due to the fact that "criminologists have the unfortunate lot of studying a phenomenon that is inherently difficult to measure" (Osgood, McMorris & Potenza, 2002: 267). As a result, data on crime are almost invariably measured with some level of error (Pepper, Petrie & Sullivan, 2011). Whereas the discipline in its early days exclusively relied on official statistics, since the 1960's criminologists have increasingly been analyzing data collected through self-report studies. These were thought to be a more valid and reliable instrument to gain insight into the real amount of crime in society because of their believed ability to identify the 'dark number'- those crimes not discovered by the police and thus not integrated in the official statistics (Short & Nye, 1957). Moreover, self-report studies offered criminologists something that official statistics never could: the possibility to test theoretically relevant correlates of delinquent behavior at the individual level, variables that are not measured in official statistics (Krohn, Thornberry, Gibson & Baldwin, 2010: 510). For these reasons, the self-report method is described as "one of the most important innovations in criminological research in the 20th century" (Thornberry & Krohn, 2000: 38).

The initial enthusiasm about the self-report method as the method par excellence to overcome the dark number and limitations of official statistics, did not last very long. A growing awareness of the limitations and biases has inspired research on the methodological aspects of the instrument, resulting in numerous improvements to the methodological and substantive aspects of the self-report instrument, gradually leading to higher quality data (Krohn, Thornberry, Bell, Lizotte & Phillips, 2012). Although it may be unrealistic to expect limitations in the self-report instrument to be completely resolved, new insights can help us to understand the various sources of bias in self-report methodology and how these affect contextual conclusions drawn from these studies (Krohn et al., 2010). In other words, systematic knowledge on this topic is highly relevant on two levels. First, methodologically, it may help improving the design of this research instrument; second, substantively, it may result in a better understanding of what the results from these surveys actually mean and how they should be interpreted.

This contribution examines the possible influence of administration mode on prevalence estimates and correlates of juvenile delinquency. More specifically, a comparison is made between estimates based on self-report data from a paper and pencil school survey and a personal self-administered mail survey. We investigate two possible ways in which these two modes of administration can impact the data, i.e. through selection effects and through measurement effects (Weisberg, 2005; Dillman et al., 2009). Respondents in both samples were drawn from the same target population, completed the same paper-and-pencil questionnaire and data were collected in the same period. This papers aims to answer

the two following research questions: (1) to what extent can differences in prevalences estimates of self-reported criminal behaviour be found between two different modes of administration? and (2) to what extent are differences in both prevalences of self-reported crime and the covariates of crime between different administration modes due to selection effects? In order to disentangle the effect of both error types on reported prevalence estimates and covariates of self-reported criminal behaviour, we will compare unmatched and matched samples of both modes of administration, as matching both samples on a selected set of sociodemographic variables would allow to control for selection effects, so that possible remaining differences could be seen as measurement effects.

Different administration modes, different results?

During the process of data collection, different types of error can occur, all of which may negatively influence the reliability and validity of the data. This in general relates to the 'total survey error' paradigm, which "refers to the accumulation of all errors that may arise in the design, collection, processing, and analysis of survey data" (Biemer, 2010: 817; Groves & Lyberg, 2010). In all, , four types of potential survey error can be identified: sampling error, coverage error, nonresponse error and measurement error (Groves, 1989; de Leeuw, Hox & Dillman, 2008). The total survey error is then to be interpreted as the sum of these four error sources in the survey design (de Leeuw et al., 2008). Sampling error occurs because a sample only contains a (small) subset of the total population, which implies that prevalence estimates will generally differ within a certain range from the 'true' population parameters. Coverage error "occurs when the operational definition of the population includes an omission, duplication, or wrongful inclusion of an element in the population" (de Leeuw et al., 2008: 17); nonresponse error relates to the fact that not all units of the sample effectively participate in the study, and measurement error occurs when participating units' observed answers are different from their 'true' answers for a certain question. Measurement error can result from processes related to the instrument, the interviewer, the respondent or data collection context (Braun, 2003; Pauwels & Pleysier, 2008).

The outcome of studies on a variety of 'sensitive' topics such as unhealthy behaviour (e.g. smoking, drinking, etc.), betting and gambling, delinquency etc., often differs considerably with regard to the reported prevalence estimates of the behaviour at stake (Kann et al., 2002; Brener et al. 2006). Irrespective of the reasons behind these differences, on which we will focus later, this aligns with findings from other studies, demonstrating the influence method and design can have on survey responses (Dillman, 2000; Tourangeau & Yan, 2007). In criminology, several studies have been conducted on the self-report survey instrument, initiated by the important work of Hirschi, Hindelang & Weiss (1979, 1981). However, most work has been done with regard to the aspect of measurement error in self-report surveys on criminal behaviour. Because of the socially undesirable character of crime most scholars have been preoccupied with discovering why people respond (dis)honestly about criminal behaviour in surveys and how the truthful responding of respondents could be increased (Junger-Tas & Haen Marshall, 1999; Thornberry & Krohn, 2000; Tourangeau & Yan, 1996; Walser & Killias, 2012). Although substantial progress has been made in the self-report methodology, both in terms of measurement and administration (Krohn et al., 2010), important questions with respect to the sources of bias in it remain unanswered. One of the central methodological issues in survey research in general and in self-report surveys of crime more specifically, is the existence of mode effects, which refers to any influence on survey responses that results from the administration mode of the survey (Jans, 2008). This may result in selection effects as well as in measurement bias. Selection effects

occur when potential respondents respond (or do not respond) in a different way to the different modes of administration (e.g. certain students may feel obliged to fill in the questionnaire at school, but choose not to participate when they receive the same questionnaire at home). This is a form of nonresponse error. Measurement effects, on the other hand, occur when different respondents would give different answers depending on the mode of administration (e.g. a student reports shoplifting in the school survey, but not in the mail survey). This is a form of measurement error. Because of these different effects of administration mode, "using more than one mode may not be advisable if the results from the two modes cannot be equated because of differences in who responds (…) and in how they respond (…)" (Sax, Gilmartin & Bryant, 2003: 412). In criminology, some studies have focused on effects of administration modes, but very specifically with regard to the question "which survey modes are most effective in eliciting self-reports of criminal or delinquent behavior?" (Kleck & Roberts, 2012: 417; Preisendörfer & Wolter, 2014). In other words, criminological research, from its preoccupation with measurement error, have studied mode effects as a potential cause of measurement error. However, before differences in reported prevalences could be defined as measurement effects, the possibility of selection effects must be taken into account and at least controlled for. When this is neglected, it remains futile to ascribe such differences in prevalence estimates to measurement effects as is generally done (Fendrich & Johnson, 2001; Kleck & Roberts, 2012)

We therefore discuss first the occurrence of selection effects, how these are affected by differential response rates and how this may impact the findings. Second, the effects of administration modes on measurement error are discussed. In a third step then, we describe how in this study we aim to disentangle both effects in order to compare the prevalence estimates and correlations of self-reported criminal behaviour in a more reliable and valid way.

a) Selection effects due to differences in nonresponse between survey modes

The nonresponse rate refers to the inability to obtain data for all sampled units, resulting from noncontact or refusal to participate. As such, a low response rate is not an inherent problem or error in the process of data collection: when nonresponse occurs processed random, this is not really a problem: although "the realized sample is smaller, resulting in larger confidence intervals around estimators, (…) the conclusions will not be biased due to nonresponse" (de Leeuw et al., 2008: 10). Nonresponse error then can be described as bias which is "a function of the amount of nonresponse and the differences between respondents and non-respondents" (Pauwels & Pleysier, 2008: 48). Bias therefore occurs when (non)respondents to a survey are not a random subgroup and therefore not 'representative' of the target population (Sheikh & Mattingly, 1981). As a result, certain groups may be underrepresented in the realized sample. This is particularly problematic if non-random differences exist with regard to the variables of interest in the study (de Leeuw et al., 2008). In other words, low response rates are not by definition equal to nonresponse bias (Dillman, 1991; Sax, Gilmartin & Bryant, 2003). If respondents' characteristics are representative of the characteristics of nonrespondents in a survey, a low response rate does not cause bias (Sax et al., 2003). High response rates in turn may reduce the risk of nonresponse bias, but are as such not a safeguard against bias (Groves & Peytcheva, 2008; Groves, 2006).

Although this specific type of bias has become more prominent due to the recent shift towards more web-based surveys (Sax et al., 2003), it is an equally important concern with regard to studies that are based on mail surveys (Van Loon et al., 2003). In general, it is found that response rates may vary by age, gender, socio-economic status, ethnic origin, marital status, etcetera. Men, young people, ethnic

minorities, people with a lower socio-economic status, non-married persons are on average less willing to participate in surveys (Kanuk & Berenson, 1975; Sheikh & Mattingly, 1981; Yammarino, Skinner & Childers, 1991). With regard to the topic of crime, the same variables found to be significantly related to differences in criminal behaviour are equally important in influencing the tendency to participate in surveys, especially in studies using mail surveys: variables as gender, age, ethnic origin, etcetera not only influence participation in survey research, but are equally related to differences in reported criminal behaviour As a consequence, the danger is that "*the very persons with the most sensitive information to report may be the least likely to report it*" (Tourangeau & Smith, 1996: 276). This suggests that the impact of nonresponse bias may be a threat to the comparability of prevalence estimates in different modes of administration, especially when the response rates of these modes strongly differ.

b) *Measurement effects due to differences in measurement error between survey modes*

In order to compare measurement effects between survey modes, survey data collection modes are often classified alongside two criteria (Kleck & Roberts, 2012). First, depending on the presence of an interviewer, a distinction is made between self-administered and interviewer-administered surveys. Second, in function of whether any computerized instruments are used, a distinction is made between paper-and-pencil versus computer assisted surveys. A consistent finding is that self-administered questionnaires result in higher levels of self-reported illicit behaviour than interviewer-administered ones (Tourangeau & Yan, 2007; Kleck & Roberts). However, it is not the physical presence of an interviewer that is important, but rather whether the respondent has to report his answers to that interviewer. A physically present interviewer unaware of the respondent's answers poses no problem (Tourangeau & Yan, 2007). Whether computer-assisted self-interviewing (CASI) leads to more accurate reporting of sensitive behaviours than traditional paper-and-pencil self-administration is still a topic of discussion. Some studies suggest that CASI, when being used among adolescents in their homes, leads to less underreporting of sensitive behaviour (Wright, Aquilino, and Supple 1998; Supple et al., 1999), while other studies have found no differences between the two modes when used in a school context (Beebe et al., 1998; Hallfors et al. 2000). These results suggest that the effect of mode varies by setting. Brener et al. (2006) systematically varied both setting and mode of administration in a randomized experiment to better understand the effect of each. They found that students were more likely to report sensitive behaviour through CASI than through paper-and-pencil instruments, but these mode effects were very weak. Also, the evidence for a mode-by-setting interaction effect as found in earlier studies was very weak. Interestingly, in light of this contribution, the strongest effect in the study of Brener et al (2006) came from the setting: students who completed a questionnaire in school were more likely to report risk behaviour than students who completed their questionnaire at home. Although not always based on an experimental design, several other studies have found similar results when comparing prevalence estimates for a variety of 'undesirable' behaviour gathered via school and in-home surveys (Fendrich & Johnson, 2001; Kann et al., 2002; Sudman, 2001). Moreover, such setting effects were most strongly identified for illegal or socially stigmatized behaviors, such as drug use and sexual promiscuous behavior; for less sensitive questions no differences were found (Brener et al., 2006; Kann et al., 2002). This 'school versus home' effect might be due to the presence of parents at home. Several studies have found that the presence of parents reduces youngsters' reporting of undesirable behaviors (Tourangeau & Yan, 2007). In contrast, it could equally be that the presence of peers during the survey completion, may stimulate respondents to 'over report' their prevalence of criminal behaviour, because of peer pressure or impression management. In certain social milieus, it

may in other words be 'cool' to report criminal behaviour, in particular among adolescents completing a questionnaire in a school context in the presence of peers (Miller, 1997).

c) The present study

Notwithstanding the growing evidence on the impact of administration modes on survey response(s), several caveats can be identified within the existing literature. First of all, in most studies, effects of the mode of administration relate to sensitive topics such as drug use, sexual behaviour, abortion, and attitudes such as homophobia, anti-Semitism, racism, etcetera. Little is known about the extent to which these findings can be generalized to other forms of criminal behaviour. In this study we therefore focus on several types of criminal behaviour such as theft, vandalism, physical violence, etcetera and aim to determine the extent to which the administration mode may impact the reported prevalence estimates, and more in particular aim to empirically disentangle selection effects from measurement effects.

Second, very few studies have compared home versus school context, which is highly relevant for studies on youth delinquency as these administration modes are most commonly used in self-report research. Generally, these studies are observational and not experimental (Fendrich & Johnson, 2001; Kann et al., 2002, Sudman, 2001; for an exception, see Brener et al 2006). As a consequence, potential forms of bias in the research design and administration mode of the compared surveys are not being controlled for, which renders contextual conclusions extremely hazardous. Enzmann (2013: 2) correctly summarizes the different conditions necessary to compare the results from different survey designs in a reliable manner: "(…) if the sampling frame is representative of the population, if there is no differential nonresponse, if the survey administration is compatible, and if underreporting is not systematically related to characteristics of the respondents, comparisons of the relative levels of crime across groups are acceptably valid – providing that the comparisons are based on data using identical or compatible questionnaires". In other words, Before a measurement effect caused by the administration mode, such as the perceived level of anonymity, could be 'held responsible' for different estimates, coverage bias and selection effects must be excluded, or being controlled for statistically. This study can control for most errors, and as a result disentangle selection from context effects, a necessary step that most studies comparing administration modes cannot of do not take into account. Third, our design is more 'realistic' than small-scale experimental research in which respondents are randomly assigned, as for example the study from Brener et al. (2006). We conducted two large-scale surveys in a way that is highly representative of most surveys into youth delinquency.

Data

In order to discern the extent to which selection effects account for differences in both the prevalences of self-reported crime and its covariates, and more specifically the role different administration modes play in this respect, other possible errors occurring in the survey design need to be minimized. The two surveys that are used to search for selection effects in this study are therefore comparable on several levels. First, with regard to potential differences in coverage error, both studies used here in principle have an comparable sampling frame, as for both surveys the target population exists of all young people attending secondary school and living in the Flemish region of Belgium. The mail

survey was sent to a representative sample of Flemish youth between 14 and 30 years old. These respondents were selected by a random sample based on the National Register, which implies that in theory every 14 to 30 year old person living in the Flemish region has an equal chance of being selected (Pleysier, 2014). In sum, 8033 persons were selected and were sent a questionnaire by mail. The Total Design Method (Dillman, 1978) was used to maximize the response. This resulted in a final response rate of 46.4% and a subsequent sample size of 3729 respondents. For the aim of this study, we selected the respondents enrolled in the 2nd and 3rd grade of secondary education. The school monitor also aimed to survey a representative sample of Flemish youth, more specifically young people from the second and third grade of secondary education. As in the mail survey, every person in this age group living in the Flemish region has equal chances of being in the sample, as in Belgium compulsory education exists until the age of 18. The survey used a two-step design, in which schools and classes were randomly selected. Within the selected classes, all pupils were asked to complete the questionnaire. In total 70 secondary schools were randomly selected (10% of all secondary schools in Flanders) using a sample matrix allowing to maximize the representativeness of the schools. If a school did not want to participate, a school with similar characteristics was drawn from a stand-by list and contacted. In sum, 128 schools had to be contacted to reach the targeted sample size of 70 secondary schools; because three schools withdraw their willingness to participate at the end, the effective number of schools was 67, which results in a global response rate of 52.3% at the school-level. Of all respondents in the sample, 89.2% filled in the questionnaire, resulting in a final sample size of 2.797 respondents. The nonresponse in this survey (i.e. 10.8%) is mainly the result of absenteeism due to sickness or classes being unable to fill in the questionnaire (because of unforeseen circumstances). However, a small (but unknown) proportion of nonresponse is caused by truancy.

As a consequence, for young people attending secondary education in the Flemish region, both sample frames in principle succeed in giving all members of the target population an equal chance of being selected. Both survey designs in other words suffer only to a very minor extent from coverage error. With regard to the mail survey, all people living in Flanders are registered in the National Register, with the exception of people living illegally in the country. As for the school survey, because of the high age of compulsory education, all young people until the age of 18 are obliged to attend secondary education. An exception to this are young people who are being taught at home and thus are not enrolled in a secondary school. However, less than 0.2% of all youngsters in the Flemish region is in this situation. In other words, the possible impact of coverage error can largely be excluded in both surveys.

Moreover, both surveys were developed and administered by the same research group, i.e. the Youth Research Platform had the same lay-out and – maybe even more important – had to a large extent an identical content. The aim of the Youth Monitor is to gain a broad insight into the living conditions, attitudes and conduct of Flemish youth. As a consequence, the monitor includes a wide range of variables relevant to the social lives of contemporary youth. Also, the similarity of the questionnaire lay-out, the questions itself and their wording, suggests that measurement error related to wording and design effects is equally not at stake in the comparison of these data collection methods. Both studies were also administered in the same period (spring 2013), so that possible seasonal or time-related influences are also kept constant to a large extent.

In all, despite the existence of some minor coverage error, based on the stark similarity of other aspects of the surveys designs of the mail and school survey, we can expect that, if differences are

found between the two studies, they can be ascribed to selection effects or to measurement effects resulting from the administration mode.

Results

Although several possible types of bias can be excluded, nonresponse bias may be a more pressing concern when comparing school survey and mail survey design. There is a large difference in the response rates in the respective studies (89.2% v. 46.4%). Although this does not imply the existence of selection effects as such, it does suggest that effects of nonresponse may have a real impact on the results in both studies, especially in the mail survey in which more than half of the contacted respondents did not return a questionnaire. First, we try to determine the extent to which the research population in both studies differ with regard to several relevant sociodemographic variables. Based on our discussion of previous findings, we control for variables such as sex, age, ethnic origin, parental professional status (as a proxy of socio-economic status) and current educational track[1]. This comparison can give an idea of the degree of homogeneity of both samples and the over- or underrepresentation of certain subgroups in both samples. This may be useful to assess the necessity to match the samples before being able to compare them in a valid way and ascribe possible differences to measurement effects.

TABLE 1. Sociodemographic variables in mail survey and school survey for young people in 2nd and 3rd grade of secondary education (with confidence intervals)

	Mail survey	School survey	Population statistics for students in secondary school
Sex (% male)	45.6 (42.2 – 48.9)	47.5 (45.7 -49.3)	51.6
Age (mean score)	16.9	16.9	/
Ethnic origin (% both parents Belgian origin)*	78.1 (75.2 – 80.8)	72.4 (70.7-74)	/
Professional status mother (% white collar, high skilled job)*	42.6 (39.2 – 46.1)	34.1 (32.3-35.9)	/
Professional status father (% white collar, high skilled job)*	56.3(52.8 - 59.8)	49.1(47.2-51)	/
Educational track (% in general secondary education)*	50.8 (47.4 - 54.2)	40.5 (38.7-42.3)	41.3

*p<.001

As the descriptives in table 1 suggest, the samples diverge on this set of sociodemographic variables. Only with regard to the 'age' variable, the two groups are not different, with the mean age of the respondents being equal in both administration modes (16.88 years on average). The other variables on the other hand, important and significant differences between both administration modes with regard to the participation of certain subgroups in the surveys. As for most of the other variables in table 1

---

[1] The Flemish educational system consists of three educational tracks. Pupils in general secondary education receive a theoretically oriented education and most of them will attend higher studies after leaving compulsory education (at the age of 18). Pupils in technical and vocational track receive a practically oriented education, preparing them for entrance into the labour market, which most of them do after leaving secondary education.

however, we do not have information on the 'true' proportion of these variables in the total population of Flemish youth, with the exception of the percentage of boys in 2nd and 3rd grade of secondary education and the distribution over the different educational tracks. For both variables, the sample of the school survey seems to align more closely to the general population, although the differential proportion of males in the school survey does not differ significantly from the percentage of males in the mail survey. This is especially so for the proportion of respondents in general secondary education, in which in reality 41.3% of all Flemish pupils are enrolled. Looking at the proportion in both studies, the sample of the school survey is almost identical with regard to this variable (with 40.5% of the sample enrolled in general education); in the mail survey in contrast students following the general educational track are overrepresented (50.8%) compared to school population statistics.

Although we only have the 'true' population scores for two of the selected variables, the comparison of respondents in both modes of administration suggest that the school survey seems to be less prone to response bias. As far as the other variables are concerned, groups known to participate less in mail surveys, such as ethnic minorities, lower SES (Brehm, 1993; Caetano, Ramisetty-Mikler, & McGrath, 2003), are indeed less present in the mail survey than in the school survey. This implies that for a meaningful comparison of prevalence estimates and correlates of criminal behaviour of both studies to be conducted, in first instance these differences in sample population must be controlled for.

Case-control matching

In order to control for this possible difference in nonresponse bias, we choose to conduct a matching technique which allows to overcome this effect of differential nonresponse. More specifically, the technique of case-control matching is used. In this statistical technique, respondents from separate samples are matched on a series of variables which may be related to nonresponse bias: sex, age, ethnic origin, professional status of mother and father (operationalized via the ISCO-classification) and current educational track. In other words, in both samples cases (individuals) are selected that are identical with regard to these six variables. To do this matching on case level, the Fuzzy extension in SPSS is used. In this study, we opt to use the mail survey sample as the 'demander' dataset and the school survey sample as the 'supplier' data set. Two reasons are relevant in this respect. First, on a practical level, the school survey sample is much larger than the mail survey (2797 respondents v. 878 respondents). Second, the comparison in table 1 suggests that the underrepresentation of certain social groups is most salient in the mail survey sample. Therefore it seems more appropriate to select a matched group from the school survey that is identical to the mail survey sample for the six selected factors.

This technique allows to compare the prevalence estimates and associations of self-reported juvenile delinquency in both unmatched and matched groups. In this latter case, we could expect to control for nonresponse bias, as the over- (or under-)representation of certain groups is accounted for, at least in both samples used here. If these differences are caused by selection effects, the differences should disappear when the matched groups are compared; if the differences in contrast remain equal after this comparison, they can be ascribed to characteristics of the administration mode. This comparison is conducted for the prevalence estimates of the delinquency items in the two surveys as well as for the associations between a set of theoretically relevant independent variables and self-reported juvenile delinquency.

Step one. Differences in prevalence estimates?

In a first step the prevalence estimates in both surveys are compared in the unmatched groups as well as in the matched groups. The comparison of the prevalence estimates can give a first idea of existing differences between the used modes of administration. In sum, nine delinquency-items were identically measured in both surveys. For each item, a 95% confidence interval is calculated to determine if the differences are significant (at the 95% level). Besides the specific crime items, we also calculated the 'general crime prevalence rate' (% of respondents reporting at least one crime) and the 'crime incidence rate' (mean number of crimes reported by the respondents). As the comparison of the prevalence estimates in table 2 show, important and significant differences for almost all self-reported criminal acts are found. While fare dodging, for example, is reported at least once by 45.6% of the respondents in the mail survey, up to 52.4% of the respondents in the school survey report to have travelled by public transport without paying at least once in the previous year. Similar conclusions can be drawn for all crimes in the surveys, with the exception of burglary. The same holds for the general crime measures, as both the prevalence and the incidence rate of crime in the school survey are higher than in the mail survey. While for example 55.3% of respondents in the mail survey report to have committed at least one crime, in the school survey up to 63.7% of respondents report at least one crime. These results therefore raise several relevant questions. At first sight, table 2 suggests that, despite the identical research design (the use of a random sample based on a solid sampling frame) and questionnaire, young people who participated in the school survey report significantly higher levels of criminal behaviour compared to respondents in the mail survey. This might lead to the conclusion that different modes of administration result in different prevalence estimates of self-reported juvenile delinquency, and that characteristics of the school survey seem to render young people more prone to report criminal behaviour.

TABLE 2. Prevalence rates (and 95% confidence intervals) of self-reported delinquency in mail survey and school survey (in percentages) in **unmatched groups**

	Mail survey	**School survey**
Fare dodging	**45.6 (42.6 - 49.5)**	**52.4 (51.0 - 54.5)**
Vandalism	**10.9 (9.1 – 13.0)**	**17.5 (16.0 – 18.8)**
Theft < 5€	**18.6 (15.8 -21.0)**	**26.1 (24.5 – 27.7)**
Theft > 5€	**4.6 (3.3 -5.9)**	**9.0 (8.2 – 10.2)**
Physical violence	**3.9 (2.7 -5.1)**	**8.3 (7.8 - 9.9)**
Drug dealing	**1.1 (0.5 -2.0)**	**4.7 (3.9 – 5.5)**
Public weapon carriage	**1.4 (0.7 -2.2)**	**5.0 (4.5 – 6.0)**
Public harassment	**3.9 (2.7 -5.1)**	**6.0 (5.2 – 6.9)**
Burglary	0.6 (0.1 -1.2)	1.2 (0.8 - 1.6)
General crime prevalence	**55.3 (51.9-58.8)**	**63.7 (61.9-65.5)**
Crime incidence rate	**1.95 (1.75-2.16)**	**3.01 (2.86-3.18)**

Repeating this analysis in the matched groups (table 3) however, reveals that some estimates change in the matched group from the school survey. For several delinquency items no or little change in prevalence estimates are observed (e.g. carrying a weapon in public theft, drug dealing) and the difference in both groups remain significant, as is shown by the 95% confidence cintervals. For fare dodging, although the prevalence estimates do not change, the difference becomes insignificant due to the larger confidence interval in the school study (caused by the lower number of respondents). For

some other items in contrast, the prevalences are lower in the matched group from the school survey. For vandalism, the prevalence rate in the matched group is 14.3% (compared to 17.5% in the unmatched group). Similar tendencies can be observed for physical violence, with a decrease from 8.3% to 5.6% and for burglary (decreasing from 1.2% to 0.6%). Although the frequencies for the first two crimes still diverge, these difference are no longer significant. When looking at the two general crime indices ('prevalence' and 'incidence'), the rates are, in line with the specific crime rates, somewhat lower in the matched group from the school survey, although especially the difference in the incidence rate remains substantial (1.95 in the mail survey versus 2.92 in the school survey sample).

TABLE 3. Prevalence rates (and 95% confidence intervals) of self-reported delinquency in mail survey and school survey **in the matched groups** (in percentages)

	Mail survey	**School survey**
Fare dodging	45.6 (42.6 - 49.5)	52.8 (49.0 – 56.9)
Vandalism	10.9 (9.1 – 13.0)	14.3 (11.8 – 17.3)
Theft $< 5€$	**18.6 (15.8 -21.0)**	**26.8 (23.3 – 30.3)**
Theft $> 5€$	**4.6 (3.3 -5.9)**	**8.5 (6.4 – 10.8)**
Physical violence	3.9 (2.7 -5.1)	5.6 (4.0 – 7.5)
Drug dealing	**1.1 (0.5 -2.0)**	**5.0 (3.4 – 6.8)**
Public weapon carriage	**1.4 (0.7 -2.2)**	**5.0 (3.4 - 6.8)**
Public harassment	3.9 (2.7 -5.1)	5.3 (3.7 – 7.2)
Burglary	0.6 (0.1 -1.2)	0.6 (0.1 – 1.2)
General crime prevalence	55.3 (51.9-58.8)	61.3 (58.0-65.0)
Crime incidence rate	**1.95 (1.75-2.16)**	**2.92 (2.62-3.27)**

Step two: differences in covariates of self-reported crime?

In a second step we test whether differences occur in the associations between a set of independent variables and the dependent variable of self-reported delinquency. The aim of these analyses is not primarily to discover new *covariates* of juvenile delinquency, but to identify possible differences between both modes of administration. To create the dependent variable the dichotomous variables of the crime-items (see table 2) are summed, resulting in a new variable with minimum score '0' (no crimes committed) to '9' (all crimes committed at least once). In order to control for measurement error –error due to differences in the operationalization of the variables – only variables measured in an identical manner in the two surveys are used.

First, a set of sociodemographic variables ('sex', 'age', 'ethnic origin' and 'professional status of father/mother') is integrated in the (linear) regression model. Second, a number of theoretical relevant, and in previous studies empirically confirmed, variables, related to juvenile delinquency, are integrated. As for the school related explanations to juvenile delinquency, three variables are used: 'current educational track' is integrated as a dummy-variable with general secondary education coded as 0 and technical and vocational secondary education coded as 1; 'relationship with teachers' is measured via a three-item Likert scale and indicates the extent to which young people evaluate the

quality of their relationship with their teachers with items such as 'teachers at school respect me'.[2] Lastly, 'frequency of truancy' is measured by an item on which respondents indicated the number of times (from '0' than 'more than 4' times) they skipped classes in the previous year. Next to these school variables, family related variables are used, most of which tap into the perceived relationship with and extent of supervision of both parents. 'Relationship with father/mother' is measured by four items with a five-item Likert scale to answer (from 'never' to 'always') and consists of items such as 'my father/mother helps me when I am having a hard time' and 'if I talk with my father/mother about my problems, he/she really helps me'.[3] 'Supervision by father/mother' is also measured by four items such as 'my father/mother knows where I go to when I go out' and 'my father/mother knows who my friends are'.[4] Finally, we integrate a dummy variable on the family situation, indicating whether parents are divorced ('1') or still living together ('0').

TABLE 4. Covariates of self-reported juvenile delinquency in mail survey and school survey **(unmatched groups)**

	Mail survey	School survey
Sociodemographic variables		
Sex (ref. cat: girl)	.080*	.108***
Age	.017	-.066***
Professional status father (ref. cat: white collar)	-.051	-.031
Professional status mother (ref. cat: white collar)	-.002	-.033
Ethnic origin (ref. cat: both parents Belgian origin)	-.033	.017
School related variables		
Technical/vocational educational track (ref. cat: general secondary education)	-.042	.053*
Relationship with teachers	-.075*	-.130***
Truancy	.268***	.295***
Family related variables		
Relationship with father	.090	-.022
Supervision by father	-.122*	-.117***
Relationship with mother	-.017	.038
Supervision by mother	-.203***	-.174***
Divorced parents (ref. cat: parents still together)	.015	.006
Adjusted R²	**.178**	**.244**

---

[2] An exploratory factor analysis reveals one underlying dimension in both surveys, with all items having factor loading between .51 and .83 in the mail survey and .61 and .83 in the school survey. The Cronbach α, as measure for the reliability of the scale, is .72 in the former survey and .78 in the latter.

[3] Both scales showed in a factor analysis to have one underlying dimension with satisfactory factor loadings for all items in both surveys (between .70 and .91). For relationship with father, the Cronbach α is .91 in the mail survey and .90 in the school survey. For relationship with mother, Cronbach α values are in both surveys .90

[4] Both scales showed in a factor analysis to have one underlying dimension with satisfactory factor loadings for all items in both surveys (between .64 and .87). For supervision by father, the Cronbach α is .87 in the mail survey and .88 in the school survey. For relationship with mother, Cronbach α values are .86 and .87 respectively.

TABLE 5. Covariates of self-reported juvenile delinquency in mail survey and school survey (**matched groups**)

	Mail survey	School survey
Sociodemographic variables		
Sex (ref. cat: girl)	.080*	.107***
Age	.017	-.031
Professional status father (ref. cat: white collar)	-.051	-.049
Professional status mother (ref. cat: white collar)	-.002	-.044
Ethnic origin (ref. cat: both parents Belgian origin)	-.033	-.007
School related variables		
Technical/vocational educational track (ref. cat: general secondary education)	-.042	.039
Relationship with teachers	-.075*	-.152***
Truancy	.268***	.294***
Family related variables		
Relationship with father	.090	.019
Supervision by father	-.122*	-.140***
Relationship with mother	-.017	-.052
Supervision by mother	-.203***	-.197***
Divorced parents (ref. cat: parents still together)	.015	-.029
Adjusted R²	**.178**	**.251**

In the unmatched groups (table 4), minor differences occur. Sex, truancy, supervision by father and supervision by mother are significantly related to self-reported delinquency, irrespective of the mode of data collection. Similarly, ethnic origin, professional status of mother and father, parental divorce and relationship with father and mother are in neither study significant correlates of self-reported delinquency. For some variables however, the relationship with delinquent behaviour does seem to differ depending on the mode of administration that is used, at least in the unmatched groups. The effect of 'age' is found to be positive and non-significant (β=.026, n.s.) in the mail survey, while in the school survey a negative and slightly significant (β=-.055, p<.05). A similar difference is found with regard to 'educational track'. While being negatively related to delinquency (β=-.047, n.s.) in the mail survey, its effect on delinquent behaviour is positive and significant in the school survey (β=.040, p<.05). In other words, in the mail survey pupils in technical/vocational education report lower levels of delinquency than general education pupils; in the school survey this relation is the opposite with pupils in general secondary education reporting lower levels of delinquent behaviour. However, the predictive value of both variables - age and educational track - is very low.

In general therefore, (with the exception of two variables) the variables that are found to be significantly related to self-reported delinquency in the unmatched and matched groups are strikingly similar. Moreover, the effect sizes appear to be almost identical in both surveys, irrespective of the matching of the samples.

Discussion and conclusion

Since the 1960's self-report studies have been the dominant methodology in assessing the amount of criminal behaviour and its main correlates (Junger-Tas & Haen-Marshall, 1999). Since its introduction in criminological research, a number of critical improvements have been made, resulting in more

reliable and valid outcomes. However, inherent bias and flaws still exist and will remain a topic of concern. In this respect, most research on the self-report survey instrument in criminology has been conducted on the topic of measurement error and differences related to the mode of administration, strongly driven by the quest to increase the amount of truthful answers on questions regarding criminal behaviour (e.g. Hindelang et al., 1981; Kleck & Roberts, 2012; Preisendorfer & Wolter, 2014). Less attention has been given to the possible influence of other types of error that can occur in the research design. Comparative studies focusing on the prevalence estimates of criminal behaviour found in different survey designs, in other words generally fail to take the possible effect of nonresponse bias into account. In order to fill this void in the literature, this paper's goal was to determine the extent to which differences in prevalence estimates of criminal behaviour in two surveys – a school survey and a mail survey – could be explained by selection effects. Most important advantage of the comparative analyses in this paper, is the high degree of similarity of the two survey designs. Both studies were (more or less) identical with regard to the target population, the extent of coverage error, the lay-out of the questionnaires and the concrete wording of the items in the questionnaires. As a consequence, coverage bias and several forms of measurement bias (e.g. wording effects) could be excluded as possible reasons for differences between the two surveys' results. However, we still need to be aware that both surveys only include a small proportion of all Flemish youth aged 14 to 18. The mere fact that a sample is used makes it necessary to understand that, despite their random selection, substantial differences between the samples can exist and the extent to which they are truly representative of the real population of Flemish youth is not fully known.

The comparison of the unmatched samples indicated that significant and important differences exist in the reported prevalences of the different types of criminal behaviour in the two surveys. For almost all behaviours, the prevalence estimates are significantly higher in the school survey, suggesting that participants in this study committed (or at least reported) more criminal behaviour than the mail survey participants. However, once both samples were matched on a selected set of relevant variables, several differences became insignificant, indicating that these differences are to a certain extent the product of diverging sample populations, and thus could be interpreted as resulting from a selection effect. However, even in the matched samples several differences remained significant. In general therefore, our findings suggest that the administration mode may influence the reported prevalence estimates of criminal behaviour both with regard to the tendency of individuals to participate in the study – selection bias - and by leading to differential tendencies to report criminal behaviour by those who do participate – measurement bias. In other words, differences in prevalence estimates of criminal behaviour between studies result from differences in the participating population as well as from differential setting or anonymity effects. Studies in which results from different surveys are compared should be aware that such comparisons should not only focus on measurement effects, but equally on selection effects as possible reasons for the explanation of diverging findings. As the main correlates of criminal behaviour are concerned, contrasting results were found, with in general very little differences between the unmatched and matched group regression analyses. This suggests that the identification of correlates or risk factors of criminal behaviour is only marginally influenced by methodological effects. As a consequence, this suggests that self-reported delinquency data are much better suited for testing criminological theories than for studying the prevalence of criminal behaviour.

Despite the use of matching techniques, self-reported criminal behaviour remains higher in the school survey than in the mail survey design. Even after controlling for possible selection effects by matching both samples on a set of relevant variables, significant differences remain. Administration modes may in other words influence the extent to which respondents are willing to admit criminal behaviour

(Lucia, Herrmann & Kilias, 2007; Pauwels & Pleysier, 2005; Kleck & Roberts, 2012). It seems clear that the school survey is better in getting young people to report criminal behaviour compared to the mail survey, being in line of previous studies on health risk behaviours in general (Fendrich & Johnson, 2001; Brener et al., 2006). The most common and popular explanation of this observation would therefore argue that the sense of anonymity and privacy is better assured in the school survey compared to the mail variant. In a mail survey, the perception of privacy safeguards may be smaller, due to the implicit or explicit presence of parents and/or other family members while completing the questionnaire (Johnston & O' Malley, 1985; Lucia et al., 2007). The social undesirability of delinquent behaviour, especially with regard to parents, may therefore result in an underreporting of this type of behaviour in a mail survey. However, diverging results might not only be due to differences in perceived response anonymity, but also due to differences in the perceived confidentiality (this involves respondents' beliefs about what happens to data after it has been collected), in the dependent measures used, the study design (random versus non-random assignment to conditions) or in the specific administration conditions (stand-alone computer versus a networked computer versus the internet) (Booth-Kewley et al., 2007).

To conclude, our results suggest that, despite the absence of golden standards with which the results from self-report research can be compared, self-reports remain a useful tool, as there are "good enough" criteria (mostly statistical then) to evaluate a self-report study. In this respect, this paper and the findings presented in it, argue for a more explicit attention to the 'Total Survey Error' paradigm ((Groves & Lyberg, 2010) in criminological research on the self-report survey. This allows and even necessitates the use of a broader and more integrative approach on errors that could possibly occur throughout the process of administering a survey on criminal behaviour. Our results indicate that criminological research could strongly benefit from more systematic attention to and studies on the different types of error. Not only can this help to ameliorate the self-report survey methodology, it can firstly be important in helping to understand better what we are measuring and how the survey design affects the results that are found. In this respect, an important lesson to be drawn is that comparing the results from different studies need to be done with caution, especially when differences between the studies are more significant than in this study, for example when samples from different populations (e.g. general youth sample vs. convicted youth delinquents) and/or different contexts (population sample vs. city sample) are compared.

Bibliography

Aquilino, W., Wright, D. & Supple, A. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use & Misuse, 35,* 845-867.

Beebe, T., Harrison, P.,Mcrae, J. Anderson, R. & Fulkerson, J. (1998). An Evaluation of Computer-Assisted Self-Interviews in a School Setting. *Public Opinion Quarterly, 62,* 623-32.

Biemer, P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*, *74,* (5), 817–848.

Brehm, J. (1993). *The phantom respondents*. Michigan: University of Michigan Press.

Braun, M. (2003). Errors in comparative survey research: an overview. In J. Harkness, F. Van de Vijver, P. Mohler (eds.). *Cross-cultural survey methods* (pp. 157-178). Hoboken: John Wiley.

Brener, N., Eaton, D., Kann, L., Grunbaum, J.A., Gross, L., Kyle, T. & Ross, J. (2006). The association of survey settings and mode with self-reported health risk behaviors among high school students, *Public Opinion Quarterly, 70,* (3), 354-374.

Caetano, R., Ramisetty-Mikler, S. & McGrath, C. (2003). Characteristics of non-respondents in a US national longitudinal survey on drinking and intimate partner violence. *Addiction*, *98,* (6), 791–797.

Cops, D. & Pleysier, S. (2014). Usual suspects, ideal victims and vice versa. The relation between youth offending and victimization and the mediating influence of risky lifestyles. *European Journal of Criminology*, 11, (3), 361-378.

De Leeuw, E., Hox, J. & Dillman, D. (2008). The cornerstones of survey research. In E. de Leeuw, J. Hox & D. Dillman (eds.). *International Handbook of Survey Methodology* (pp. 1-17). London: Routledge.

De Leeuw, E. (2008). Choosing the method of data collection. In E. de Leeuw, J. Hox & D. Dillman (eds.). *International Handbook of Survey Methodology* (pp. 114-135). London: Routledge.

Dillman, D. (1978). *Mail and telephone surveys: the total design method*. New York: Wiley.

Dillman, D. (2000). *Mail and internet surveys: The tailored design method*. New York: Wiley.

Enzmann, D. (2013). The Impact of Questionnaire Design on Prevalence and Incidence Rate of Self-Reported Delinquency: Results of an Experiment Modifying the ISRD-2 Questionnaire. *Journal of Contemporary Criminal Justice*, *29*(1), 147–177.

Fendrich, M. & Johnson, T. (2001). Examining prevalence differences in three national surveys of youth: impact of consent procedures, mode, and editing rules. *Journal of Drug Issues, 31,* (3), 615-642.

Groves, R. (1989). *Survey errors and survey costs*. New York: John Wiley.

Groves, R. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70,* (5), 646-675.

Groves, R, & Lyberg, L. (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly*, *74,* (5), 849–879.

Groves, R. & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. A meta-analysis. *Public Opinion Quarterly, 72*, (2), 167-189.

Hagan, F. (2006). *Research methods in criminal justice and criminology*. Boston: Allyn and Bacon.

Hallfors, D. Khatapoush, S.,Kadushin, C. Watson, K. & Saxe, L. (2000). A Comparison of Paper vs. Computer-Assisted Self Interview for School Alcohol, Tobacco, and Other Drug Surveys. *Evaluation and Program Planning, 23,* 149-55.

Junger-Tas, J., & Haen-Marshall, I. (1999). The self-report methodology in crime research. *Crime and Justice*, (25), 291–367.

Junger-Tas, J., Haen Marshall, I., Enzmann, D., Killias, M. & Gruszcynska, B. (2010) (eds.). *Juvenile Delinquency in Europe and beyond. Results of the second International Self-Report Delinquency Study.* New York: Springer.

Kann, L., Brener, N., Warren, C., Collins, J. & Giovino, G. (2002). An assessment of the effect of data collection setting on the prevalence of health risk behaviors among adolescents. *Journal of Adolescent Health, 31,* 327-335.

Kanuk, L., & Berenson, C. (1975). Mail surveys and response rates: A literature review. *Journal of Marketing Research*, 440–453.

Kleck, G. & Roberts, K. (2012). What surveys methods are most effective in eliciting self-reports of criminal or delinquent behavior? In L. Gideon (ed.). *Handbook of survey methodology for the social sciences* (pp. 417-436). New York: Springer.

Krohn, M., Thornberry, T., Gibson, C. & Baldwin, J. (2010). The development and impact of self-report measures of crime and delinquency, *Journal of Quantitative Criminology, 26*, 509-525.

Krohn, M., Thornberry, T., Bell, K., Lizotte, A. & Philips, M. (2012). Self-report surveys within longitudinal panel designs. In D. Gadd, S. Karstedt & S. Messner (eds.). *The Sage Handbook of Criminological research methods* (pp. 23-35). London: Sage.

Krohn, M., Lizotte, A., Phillips, M., Thornberry, T. & Bell, K. (2013). Explaining systematic bias in self-reported measures: factors that affect the under- and over-reporting of self-reported arrests. *Justice Quarterly, 30,* (3), 501-528.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality and Quantity, 47,* 2025-2047.

Kuha, J. & Jackson, J. (2014). The item count method for sensitive survey questions: modelling criminal behaviour. *Applied Statistics*, published online, 1-21.

Lynch, J. & Addington, L. (2010). Identifying and addressing response errors in self-report surveys. In A. Piquero & D. Weisburd (eds.). *Handbook of quantitative criminology* (pp. 251-272). New York: Springer.

Macera, C., Jackson, K., Davis, D., Kronenfeld, J. & Blair, S. (1990). Patterns of nonresponse to a mail survey, *Journal of Clinical Epidemiology, 43,* (12), 1427-1430.

Miller, P. (1997). Is "up" right? The National Household Survey on Drug Abuse (the polls review). *Public Opinion Quarterly, 64,* 627-641.

Nettler, G. 1978. *Explaining crime.* New York: McGraw-Hill.

Pauwels, L. & Pleysier, S. (2008). Crime victims and insecurity surveys in Belgium and the Netherlands. In R. Zauberman (ed.). *Victimisation and insecurity in Europe. A review of surveys and their use* (pp. 39-64). Brussels: VUBPress.

Pepper, J., Petrie, C. & Sullivan, S. (2011). Measurement error in criminal justice data. In A. Piquero & D. Weisburd (eds.). *Handbook of Quantitative Criminology* (pp. 353-374). New York: Springer.

Percy, A., McAlister, S., Higgins, K., McCrystal, P. & Thornoton, M. (2005). Response Consistency in Young Adolescents' Drug Use Self-Reports: A Recanting Rate Analysis. *Addiction, 100,* 189-96.

Pleysier S. (2014). 'Ondanks beperkt zicht'. De context van 'jongeren in cijfers en letters'. In L. Bradt, S. Pleysier, J. Put, J. Siongers & B. Spruyt (eds.) *Jongeren in cijfers en letters. Bevindingen uit de JOP-monitor 3 en de JOP-schoolmonitor 2013 (pp. 19-39).* Acco: Leuven.

Preisendörfer, P. & Wolter, F. (2014). Who is telling the truth? A validation study on determinants of response behavior in surveys, *Public Opinion Quarterly*,

Osgood, W., McMorris, B. & Potenza, M. (2002). Analyzing multiple-item measures of crime and deviance I: item response theory scaling, *Journal of Quantitative Criminology, 18,* 267-296.

Sax, L., Gilmartin, S. & Bryant, A. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education, 44,* (4), 409-432.

Sheikh, K., & Mattingly, S. (1981). Investigating nonresponse bias in mail surveys. *Journal of Epidemiology and Community Health*, *35,* (4), 293–296.

Sudman, S. (2001). Examining Substance Abuse Data Collection Methodologies. Journal of Drug Issues 31, 695-716

Supple, A., Aquilino, W. & Wright, D. (1999). Collecting sensitive self-report data with laptop computers. *Journal of Research on Adolescence, 9,* (4), 467-488.

Thornberry, T., & Krohn, M. (2000). The self-report method for measuring delinquency and crime. *Criminal Justice*, *4*(1), 33–83.

Tourangeau, R. & Smith, T. (1996). Asking sensitive questions. The impact of data collection, question format en question context. *Public Opinion Quarterly, 60,* 275-304.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883.

Van Loon, J., Tijhuis, M., Picavet, S., Surtees, P. & Ormel, J. (2003). Survey nonresponse in the Netherlands: effects on prevalence estimates and associations, *Ann Epidemiol, 13*, 105-110.

Walser, S., & Killias, M. (2012). Who should supervise students during self-report interviews? A controlled experiment on response behavior in online questionnaires. *Journal of Experimental Criminology*, *8*(1), 17–28.

Yammarino, F., Skinner, S., & Childers, T. (1991). Understanding mail survey response behavior a meta-analysis. *Public Opinion Quarterly*, *55,*(4), 613–639.