

Who should supervise students during self-report interviews? A controlled experiment on response behavior in online questionnaires

Simone Walser · Martin Killias

Published online: 3 August 2011
© Springer Science+Business Media B.V. 2011

Abstract

Objectives This study tests whether juveniles' responses on sensitive topics such as self-reported delinquency, victimization, and substance use are comparable when teachers versus external persons supervise students while filling out online questionnaires.

Methods Eighty classes with 1,197 students (9th grade) in eastern Switzerland were randomly assigned (at the class level) to supervision either by their teacher or by an external person (i.e., researcher). Students filled out online questionnaires about self-reported delinquency, victimization, and substance use in the classroom while being supervised either by their teacher or by an external person. Prevalence rates were compared using Chi-square tests.

Results Only three out of 57 comparisons show significantly different outcomes. Whenever differences are found, and contrary to our expectation, sensitive experiences are more often admitted when students are supervised by their teacher. Effect sizes do not exceed 0.6, and are thus all below the limit of a small effect.

Conclusions Using online questionnaires with teachers as supervisors may not affect validity while making surveys less expensive and intrusive.

Keywords Internet surveys · Interview methods · Research methods · School-based surveys · Self-reported juvenile delinquency

Introduction

In the area of self-administered self-report surveys based on students in the classroom, it has tacitly been assumed that data validity is greatly enhanced if

S. Walser (✉) · M. Killias
Institute of Criminology, University of Zurich, Raemistrasse 74/39, 8001 Zurich, Switzerland
e-mail: simone.walser@rwi.uzh.ch

M. Killias
e-mail: martin.killias@rwi.uzh.ch

students are being supervised by external researchers rather than by teachers. Whether this assumption is valid has been tested only relatively rarely. In the following section, we will review the current state of knowledge. As will be shown, the few available tests all concern surveys with paper-and-pencil questionnaires. Over the last years, online questionnaires have become increasingly popular, however, and it is not known whether observations made with paper-and-pencil questionnaires are still valid. In the main part of this paper, we shall describe a randomized controlled trial through which the role of supervisors on response behavior among students while filling out online questionnaires has been tested. The findings suggest that supervision by teachers yields equally valid data than supervision by an external person. In the final section, we shall discuss the implications of this finding. As will be shown, the savings that online questionnaires without external supervisors allow will greatly contribute to make self-reported surveys far more feasible even in countries or under circumstances where research budgets are tight.

The current state of knowledge

Surveys about sensitive information, i.e., information about illegal or embarrassing behavior such as involvement in delinquency, drug use, or sexual behavior, have become increasingly popular over the last decades (Tourangeau and Smith 1996). The vast majority of these surveys rely on self-reports, thus, issues about validity of responses have been a subject of debate and research since the first times they were used (Hindelang et al. 1979). This aspect has become increasingly important given the increased reliance on data from these surveys in determining appropriate public policy (Gribble et al. 1999). Researchers agree on the importance of the perceived level of privacy or confidentiality, provided by the method of data collection, on respondents' willingness to reveal illegal or socially undesirable behavior (see e.g., Aquilino 1994; Brener et al. 2003; Turner et al. 1998).

One important aspect concerning anonymity during data collection is the way questions are asked and the way responses are collected. While interviewer-administered questionnaires (IAQs) involve an interviewer posing the questions and recoding the answers, self-administered questionnaires (SAQs) are read and filled out directly by the respondent. Working under the assumption that higher prevalence rates are more accurate than lower ones in reports of illegal or embarrassing behavior (Gans and Brindis 1995; Turner et al. 1998), SAQs are generally more valid than IAQs (see e.g., Aquilino 1994; Brener et al. 2003; Tourangeau and Smith 1996). By not having an interviewer writing down the answers, the respondent does not have to reveal embarrassing personal information directly to another person and, thus, reporting bias can be reduced (Thornberry and Krohn 2000; Trapl 2007).

Another aspect possibly influencing anonymity and privacy during interviews is the setting in which interviews take place. When adolescents are investigated, they are usually interviewed at school (i.e., during classes). In comparison to interviews at home, this procedure is time-saving because an entire school class can be interviewed together and no time-consuming home visits are required. Besides this administrative advantage, responses given in the classroom are no less valid than

those given at home (Kann et al. 2002). However, when interviews are conducted in the classroom, it must be considered that elements of the environment may also impact perception of privacy (Trapl 2007). Anonymity can be increased if, for example, desks are moved far apart or teachers are discouraged from walking among the students while they are completing the questionnaire (Beebe et al. 1998; Gans and Brindis 1995).

In recent years, online questionnaires, which are filled out at a computer, have become increasingly popular as an alternative to the traditional paper-and-pencil form. Computerized approaches offer multiple advantages over paper-and-pencil surveys. Expenses can be saved as no questionnaires need to be printed and no responses given on paper need to be coded. Additionally, transcribing responses from paper into electronic format always implies risks of typing errors. With online questionnaires, respondents face less difficulties (and risks of getting confused) in filling out complex questionnaires (i.e., with branching or skip patterns) since computer programs with sophisticated filters automatically skip unnecessary follow-up questions and lead directly to the next relevant question. Further, computer questionnaires decrease the incidence of missing data (i.e., item non-response) and improve the speed of data processing (Hallfors et al. 2000). In addition to these methodological advantages, the available evidence shows that responses, including responses on sensitive topics, collected in the classroom through computer questionnaires are comparable to those gathered through paper-and-pencil forms (Brener et al. 2003; Hallfors et al. 2000; Lucia et al. 2007). Some researchers even report more self-disclosure concerning sensitive behaviors with computerized than with traditional paper-and-pencil forms (Vereecken and Maes 2006; Weisband and Kiesler 1996). Results are similar when interviews are conducted at home (Turner et al. 1998; Wright et al. 1998). Thus, it seems that computer questionnaires can figure as a valid alternative to traditional paper-and-pencil formats.

With IAQs, interviewers must be trained and instructed carefully about how to carry out the interviews, thus, they are usually engaged and paid from the research institute. With SAQs, the interviewer does not directly interact with respondents given that questionnaires are read and filled out without additional assistance. The function of the interviewer, then, is that of a supervisor, i.e., to make sure that no disturbances occur during interviews and to be the contact for possible comprehension problems. Thus, the question arises if other, more available, persons could figure as supervisors instead of external persons. If students are interviewed at school, it would be feasible to let a teacher supervise them. Particularly in larger countries, sending researchers to remote areas has serious cost implications that have so far prevented many countries from conducting surveys on juvenile delinquency and similar issues. The use of online questionnaires combined with teacher supervision, allows increasing samples without additional field costs (as long as schools are not paid for their participation). Larger samples, in turn, allow the investigation of rare events like the commission of severe offences or the consumption of hard drugs.

It is surprising that, given these advantages for supervision by teachers, most surveys conducted in school classes are still being supervised by an external person (e.g., Beebe et al. 1998; Hallfors et al. 2000; Lucia et al. 2007). To our knowledge, the validity of responses has never been compared experimentally when an external

person versus a teacher supervises students during online interviews. It seems to have been tacitly assumed that the presence of a teacher (who is well known by the students and act as ‘authority figure’) decreases students’ willingness to admit sensitive experiences compared to when they are supervised by an external (i.e., unknown) person. The only study on this topic seems to be the one conducted in Finland (Kivivuori and Salmi 2011) where 26 classes (9th grade) were assigned either to teacher or to researcher supervision. In the condition where researchers, dispatched to the classroom from the research institute, supervised students, significantly higher rates of self-reported delinquency were observed for destruction of property and drug use but not for violence and theft. Higher victimization rates under researcher supervision have been observed only for theft. All in all, the authors conclude that there appears to be no critical risks on choosing between researcher and teacher supervision. However, this study was conducted with paper-and-pencil questionnaires and therefore does not allow conclusions with regard to online questionnaires. Furthermore, the study suffers from some methodological limitation such as a rather small sample size (26 classes with a total of 482 respondents), group level randomization and a matching process which was formed on the basis of teachers’ assessments of behavioral and ‘criminal’ problems in their classes.

Our study tries to fill this gap by looking at the validity of sensitive data collected online when students are supervised by their teacher or, alternatively, by an external and, thus, probably more ‘anonymous’ researcher.

Methods

Data set

An opportunity to test this issue was provided by a survey on self-reported delinquency among 9th grade students (aged between 15 and 16 years) enrolled in public junior high schools in the Canton of St. Gallen (eastern Switzerland). High schools are subdivided into four levels according to students’ intellectual performance (i.e., ‘Gymnasium’, the highest level, ‘Sekundarschule’, ‘Realschule’, and ‘Kleinklasse’, the lowest level). Private schools are attended by less than 2% of all students and were not included. Special schools for students with physical or mental disabilities were also excluded because these students might need assistance in filling out a questionnaire so that anonymity could not be guaranteed.

Out of a total of 376 9th grade classes in the Canton of St. Gallen, 338 agreed to participate. For the experiment presented here, 80 classes were randomly selected and randomly assigned to the experimental and the control condition (i.e., 40 classes per condition). The proportion of the four school levels in the 80 classes (and in both conditions) was the same as in the entire sample, resulting in four classes of the ‘Gymnasium’, 17 classes of the ‘Sekundarschule’, 15 classes of the ‘Realschule’, and four classes of the ‘Kleinklasse’ in each condition. The number of students was comparable in both groups with 669 students in the experimental (i.e., supervision by an external person) and 672 students in the control group (i.e., supervision by a teacher). The number of males and females did not differ significantly across the

groups. Only two students of the control group and none of the experimental group refused to participate. For several reasons (such as illness), 7.6% of students in the experimental and 6.5% in the control group were absent from school the day the survey was conducted. Due to computer crashes during completion, 15 questionnaires were lost in the experimental and 25 in the control group. After removal of a few questionnaires with obviously invalid responses (five in the experimental group and two in the control group), 598 questionnaires in the experimental group and 599 in the control group remained available for analyses.

Chi-square tests confirmed that students did not differ significantly across groups with respect to gender (50.8% females in the experimental versus 48.7% in the control group) and age (mean 15.28 years in the experimental and 15.25 years in the control group). In sum, the sample was based on a cluster randomized trial (with classes as clusters) with data on the students' level. Classes were randomly assigned to two groups (supervision by teacher or researcher), which were comparable with regard to size, gender, age, and school level.

Data collection

Data collection lasted from January to April 2008. Teachers of the experimental group were informed that an external person (i.e., a senior student of our institute) was going to supervise their students during online interviews. Teachers of the control group were informed that they should supervise their students during interviews. An information sheet was handed out to parents by teachers. Passive consent was applied, but none of the parents refused their child to participate. The questionnaire was sent in advance to all teachers of the control group with the instruction to report any possible problems of understanding. Pre-tests had shown that the questions were self-explanatory for students (and thus for teachers as well). Teachers were also informed about how to organize the online interviews. Within the limits of the fieldwork period (January to April 2008), teachers were free to choose the time of the interview session. For the experimental group, the external supervisors were invited to contact the teachers of their classes and to agree with them on a date for the interview session.

Computers and the Internet were available in all 9th grade classes in the Canton of St. Gallen. In the control group, the teacher supervised the class during the test. In the experimental group, students were supervised by a researcher dispatched by our institute, and the teacher was asked to leave the room. To guarantee maximum anonymity for students, supervisors were advised not to walk around or to watch the computer screens while students were filling out the questionnaire. Whenever students had a problem, they were told to see the supervisor at his desk. Supervisors were asked to stress the anonymity of replies and that neither parents, the police, the school, nor anybody else would have access to their responses.

The questionnaire was presented with computer-assisted-self-interviewing (CASI) using a program from NETQ (www.netq.ch), which allows to fill out questionnaires online and to transfer data electronically and automatically to different types of data files such as SPSS (Statistical Package for the Social Sciences) or Excel. The questionnaire could be used with a PC or a Mac. Questions were presented one by one, i.e., the next relevant question did not appear on the screen until the answer of

the preceding question was confirmed. Supervision by teachers or researchers did not affect the time needed to complete the questionnaire (t test). On average, 26 min was needed (in both groups) to fill out the questionnaires; less than 2% of all students needed more than an entire lesson of 45 min.

Measures

The questionnaire included questions concerning self-reported delinquency, victimization, and substance use. Furthermore, it contained variables about social background, family, school, leisure time, and routine activities. In the present context, only results regarding sensitive issues that might be affected by the mode of supervision, i.e., self-reported delinquency, victimization, and substance use, will be presented. The self-reported delinquency items included four violent offences (assault, group fight, robbery, and sexual assault), four property offences (burglary, shoplifting, bicycle theft, and other theft) and three other offences (vandalism, carrying a weapon, and drug dealing). All 11 offences were aggregated to an index (i.e., any delinquency) to classify those respondents who have committed at least one of the above-mentioned offences. Items on victimization included assault, robbery, and sexual assault. As with self-reported delinquency, all three offences were collapsed into a global index (any victimization) to see whether respondents had been victimized at least once of any of these offences or not at all. Substance use included beer/wine, strong alcohol (and an index of any alcohol), cannabis, cocaine, heroine, ecstasy/amphetamine/speed, and LSD/psilocybin, mushrooms (and an index of any illicit drug). Lifetime ('Have you ever...?') and last year ('...over the last 12 months?') prevalence was collected for each behavior (self-reported delinquency, victimization, and substance use). Response options were, in the first place, dichotomous ('yes' or 'no'), but it was also possible to answer 'I do not want to respond'. In this way, non-response could also be assessed.

Results

We compared lifetime and last-year prevalence rates of self-reported delinquency, victimization, and substance use as well as rates of missing values (i.e., item non-response) among the 598 students in the experimental group (supervision by researcher) and the 599 students in the control group (supervision by teacher). Given the dichotomous nature of the dependent variables, Pearson's Chi-square test was used for all analyses (with a p value of 0.05 was considered as significant). Furthermore, effect sizes (Cramer's V) are provided.

Whenever randomization is carried out on the class level (with data on the students' level), analysis should take into account possible cluster effects (i.e., the fact that individuals within clusters tend to be more similar than individuals in different clusters). Because data protection procedures did not allow collecting information on the class respondents were attending, students could not be allocated to classes. Therefore, no cluster effects could be controlled for. This limitation can lead to p values that are too small (Bennett et al. 2002; Bland 2004). Given that the hypothesis to be tested suggests no differences across groups, this limitation would

favor a conservative outcome, i.e., it increases the odds of finding a significant difference across forms of supervision. With regard to this uncertainty, the measures of association (Cramer's V) that are indicated in the tables may merit more attention than the mere measures of significance. Furthermore, as the sizes of classes (clusters) are not equal, no power analysis can be provided.

No significant differences in lifetime prevalence rates of self-reported offending are found for the 11 single offences and for any delinquency (Table 1). Thus, self-reports of having ever committed any of the 11 offences do not differ between supervision by the teacher (control group) and an external person (experimental group). Last-year prevalence rates of self-reported offending are also similar in both groups (Table 1). Only group fights and shoplifting over the previous 12 months are admitted at a significantly ($p < 0.05$) higher rate if the interviews were supervised by the teacher. In these two cases, Cramer's V is 0.06, whereas in all other comparisons it is consistently smaller (between 0.00 and 0.04).

No significant differences between the two groups were found for lifetime and last-year prevalence rates regarding victimization of any of the three offences nor for victimization overall (Table 2). Supervision by the teacher or an external person does not affect self-reported substance use either (ever or during the last year, Table 3). In both tables, Cramer's V never exceeds 0.05 and most of the times remains close to zero.

The questionnaire allowed for all items about self-reported delinquency, victimization, and substance use to answer 'I do not want to respond'. Given that less than 3% opted for this answer at any item on average, non-response will be

Table 1 Lifetime and last-year prevalence rates of self-reported delinquency by supervision mode (in %)

Offenders (%)	Lifetime prevalence				Last-year prevalence			
	Supervision by				Supervision by			
	Researcher (<i>n</i> =598)	Teacher (<i>n</i> =599)	χ^2 (df=1)	Cramer's V	Researcher (<i>n</i> =598)	Teacher (<i>n</i> =599)	χ^2 (df=1)	Cramer's V
Assault	19.4	18.4	0.23	0.01	14.1	13.9	0.01	0.00
Group fight	17.4	18.5	0.25	0.02	12.9	17.3	4.37*	0.06
Robbery	4.3	4.6	0.07	0.01	3.6	3.9	0.08	0.01
Sexual assault	1.0	0.9	0.11	0.01	1.0	0.9	0.11	0.01
Burglary	4.3	5.4	0.88	0.03	3.4	4.4	0.77	0.03
Shoplifting	36.8	40.6	1.79	0.04	23.6	29.2	4.71*	0.06
Bicycle theft	16.3	18.8	1.29	0.03	13.8	15.1	0.38	0.02
Other theft	33.4	29.8	1.77	0.04	22.4	23.1	0.06	0.01
Vandalism	22.7	23.4	0.09	0.01	19.4	20.5	0.25	0.02
Carrying a weapon	19.4	18.2	0.30	0.02	16.6	16.3	0.02	0.00
Drug dealing	9.8	11.8	1.28	0.03	8.4	10.4	1.31	0.03
Any delinquency ^a	65.4	63.8	0.37	0.02	49.5	53.2	1.63	0.04

*Significant with $p < 0.05$

^a Assault, group fight, robbery, sexual assault, burglary, shoplifting, bicycle theft, other theft, vandalism, carrying a weapon, and drug dealing

Table 2 Lifetime and last-year prevalence rates of victimization by supervision mode (in %)

Victims (%)	Lifetime prevalence				Last-year prevalence			
	Supervision by				Supervision by			
	Researcher (n=598)	Teacher (n=599)	χ^2 (df=1)	Cramer's V	Researcher (n=598)	Teacher (n=599)	χ^2 (df=1)	Cramer's V
Assault	14.6	15.4	0.15	0.01	10.6	10.8	0.02	0.00
Robbery	15.6	15.5	0.00	0.00	11.6	12.4	0.17	0.01
Sexual assault	6.8	5.4	1.05	0.03	4.6	3.2	1.56	0.04
Any victimization ^a	26.0	27.7	0.47	0.02	20.1	20.4	0.01	0.00

NOTE: All Chi-square statistics are non-significant with $p > 0.05$

^a Assault, robbery, and sexual assault

analyzed only globally for all items of self-reported delinquency, victimization, consumption of alcohol, and use of any illicit drugs. As it turned out, non-response does not differ significantly across groups (Table 4), except for self-reported delinquency (ever) where it is more frequent if students were supervised by an external person rather than by their teacher ($p < 0.01$). Even in this case, however, Cramer's V is only 0.02.

Using a significance level of 0.05, we accept to find by chance χ^2 five significant comparisons out of 100. With three significant comparisons out of 57, the proportion of significant outcomes is slightly above 5%. However, since cluster effects have not been taken into account, the number of significant differences may be too high. Cramer's V, as indicated for every comparison (see Tables 1, 2, 3, and 4), provides a

Table 3 Lifetime and last-year prevalence rates of substance use by supervision mode (in %)

Users (%)	Lifetime prevalence				Last-year prevalence			
	Supervision by				Supervision by			
	Researcher (n=598)	Teacher (n=599)	χ^2 (df=1)	Cramer's V	Researcher (n=598)	Teacher (n=599)	χ^2 (df=1)	Cramer's V
Beer/wine	83.9	83.7	0.00	0.00	80.8	80.7	0.00	0.00
Strong alcohol	67.7	70.0	0.69	0.02	63.3	65.4	0.58	0.02
Any alcohol ^a	85.0	84.8	0.01	0.00	81.5	81.6	0.00	0.00
Cannabis	31.3	32.3	0.15	0.01	29.3	28.2	0.15	0.01
Cocaine	4.4	3.9	0.18	0.01	4.0	3.0	0.87	0.03
Heroin	1.9	1.2	0.91	0.03	1.7	0.8	1.69	0.04
Ecstasy/amphetamine/speed	4.2	2.5	2.52	0.05	3.5	2.2	1.90	0.04
LSD/psilocybin, mushrooms	2.2	2.7	0.32	0.02	2.2	2.2	0.00	0.00
Any drug ^b	31.5	33.2	0.38	0.02	29.9	28.7	0.20	0.01

NOTE: All Chi-square statistics are non-significant with $p > 0.05$

^a Beer/wine and strong alcohol

^b Cannabis, cocaine, heroin, ecstasy/amphetamine/speed, and LSD/psilocybin, mushrooms

Table 4 Item non-response (missing values) of self-reported delinquency, victimization, and substance use by supervision mode (in %)

Item non-response (%)	Lifetime prevalence				Last-year prevalence			
	Supervision by							
	Researcher (n=598)	Teacher (n=599)	χ^2 (df=1)	Cramer's V	Researcher (n=598)	Teacher (n=599)	χ^2 (df=1)	Cramer's V
Self-reported delinquency ^a	2.6	1.9	7.59 **	0.02	0.8	0.7	0.91	0.01
Victimization ^b	1.6	0.9	2.74	0.03	0.6	0.5	0.22	0.01
Alcohol use ^c	1.5	0.8	3.05	0.04	1.4	1.0	0.62	0.02
Drug use ^d	0.9	1.1	0.41	0.01	- ^e	- ^e	- ^e	- ^e

**Significant with $p < 0.01$

^a Assault, group fight, robbery, sexual assault, burglary, shoplifting, bicycle theft, other theft, vandalism, carrying a weapon, and drug dealing

^b Assault, robbery, and sexual assault

^c Beer/wine and strong alcohol

^d Cannabis, cocaine, heroine, ecstasy/amphetamine/speed and LSD/psilocybin, mushrooms

^e No statistical analysis allowed due to small numbers

useful effect size index for Pearson's Chi-square tests (Cohen 1988: 223). Effect sizes range from 0.00 to 0.06 with an average effect size of 0.02, thus, considerably below 0.1, the value which Cohen (1988: 227) considers the critical limit for a small effect. In sum, analyses of statistical significance as well as effect sizes indicate that there is at most a marginal effect of the supervising person (teacher versus external, unknown person) on adolescents' response behavior concerning self-reported delinquency, victimization, and substance use. Surprisingly, and even worse for the hypothesis of presumably higher rates of reporting under supervision by a researcher, the three significant comparisons show prevalence rates to be higher and non-response to be lower when students are supervised by their teacher.

Discussion

In the present study, adolescents' responses on sensitive data, collected with an online questionnaire, are compared under two conditions, namely, when students are supervised by their teacher or by an external (i.e., unknown) person. Overall, lifetime and last-year prevalence rates as well as missing values (i.e., non-response) of self-reported delinquency, victimization, and substance use are similar for the two conditions. Only three out of 57 comparisons differ significantly and, contrary to expectation, in these three cases, self-reported behavior (group fights and shoplifting) is reported at higher rates and item non-response of all delinquency items taken together is smaller when students are supervised by their teacher.

Contrary to our results, Kivivuori and Salmi (2011) report rather higher rates of self-reported delinquency and substance use when adolescents were supervised by an external person. Their results may be due to the fact that they used paper-and-

pencil questionnaires. Compared to such forms, online questionnaires probably offer more anonymity for several reasons: (1) Answers given by students cannot be read by third parties unless the observer stands immediately behind the screen (and the student)—what supervisors were explicitly advised not to do in our study. Pages of printed forms may also be readable even from some distance; (2) Online programs can show one question after the other. Even if standing behind the student, an observer thus would only see one answer. On a paper sheet, however, several questions are usually listed on one page, and responses remain visible until the entire page is filled out. Since paper-and-pencil questionnaires do not allow offering full anonymity, adolescents may be more concerned about the consequences from their teacher than those from an external and unknown person whom they will probably never see again.

For future self-report studies, the findings have several important implications. Most notably, the cost of self-report studies can be substantially reduced if no researcher has to be dispatched to supervise classes located in remote areas. With online interviews supervised by teachers, it will be possible to conduct such surveys at far lower costs. In the case of our study, about 85,000 Swiss francs (or US \$103,400) would have been needed to dispatch researchers to all classes. Thus, supervision costs would have been by far the most important position in the budget of the entire research project. With online questionnaires and internal supervision by teachers, self-report studies can be conducted with almost no field costs whatever the size of the sample. In this way, they will become feasible even in countries with scarce research resources—provided that computers with Internet access are available in schools. It can be expected that these conditions will be met by an increasing number of countries in the near future.

Supervision by the teacher is also more attractive for teachers and schools since no appointment with a researcher needs to be set for the time of the interviews. Teachers are free to choose the moment of the interviews within a period of several weeks or months and can thus use the survey to fill an ‘empty’ cell in their schedule. This, along with the attraction of online questionnaires, will likely increase participation among school principals, teachers, and students. Higher participation rates increase the internal validity of self-report studies. Apparently, teachers were able to answer students’ eventual questions, thanks to the materials that were made available to them way in advance and to the self-explaining character of the questionnaire. At least, no problems at this level affected response behavior.

A drawback may come from the fact that classes rather than individuals were randomized. This may have reduced statistical power. As no information regarding the place of residence or the school or even class was asked in order to provide full anonymity to our subjects, individuals cannot be allocated to clusters (i.e., classes) and cluster effects cannot be controlled for. This involves the risk that some actually non-significant differences may appear to be significant. This leads to the conclusion that if we had controlled for cluster effects, we probably would have found even less significant differences. In the end, however, our findings have to be taken with caution and further research on this topic is needed.

In conclusion, online questionnaires and supervision by teachers may help to facilitate interviewing students on sensitive issues at schools and contribute to the

diffusion of such data collection initiatives in areas of the world where such methods have, so far, been considered as unaffordable. For the field of delinquency research and comparative criminology, this innovation may therefore show great potential.

Acknowledgements This survey was mandated by the Government of the Canton of St. Gallen, Switzerland (Department of Security & Justice and Department of Education) and carried out by the Institute of Criminology, University of Zurich, Switzerland.

References

- Aquilino, W. S. (1994). Interview mode effects in surveys of drug and alcohol use: A field experiment. *Public Opinion Quarterly*, *58*(2), 210–240.
- Beebe, T. J., Harrison, P. A., McRae, J. A., Anderson, R. E., & Fulkerson, J. A. (1998). An evaluation of computer-assisted self-interviews in a school setting. *Public Opinion Quarterly*, *62*(4), 623–632.
- Bennett, S., Parpia, T., Hayes, R., & Cousens, S. (2002). Methods for the analysis of incidence rates in cluster randomized trials. *International Journal of Epidemiology*, *31*(4), 839–846.
- Bland, J. M. (2004). Cluster randomised trials in the medical literature: Two bibliometric surveys. *BMC Medical Research Methodology*, *4*(1), 21–26.
- Brener, N. D., Billy, J. O. G., & Grady, W. R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health*, *33*(6), 436–457.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum.
- Gans, J. E., & Brindis, C. D. (1995). Choice of research setting in understanding adolescent health problems. *Journal of Adolescent Health*, *17*(5), 306–313.
- Gribble, J. N., Miller, H. G., Rogers, S. M., & Turner, C. F. (1999). Interview mode and measurement of sexual behaviors: Methodological issues. *Journal of Sex Research*, *36*(1), 16–17.
- 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*(2), 149–155.
- Hindelang, M. J., Hirschi, T., & Weis, J. G. (1979). Correlates of delinquency: The illusion of discrepancy between self-report and official measures. *American Sociological Review*, *44*(6), 995–1014.
- Kann, L., Brener, N. D., Warren, C. W., Collins, J. L., & Giovino, G. A. (2002). An assessment of the effect of data collection setting on the prevalence of health risk behaviors among adolescents. *Journal of Adolescent Health*, *31*(4), 327–335.
- Kivivuori, J. & Salmi, V. (2011). *Supervision mode effects in school-based delinquency and victimization survey: Preliminary test* (Research Brief No. 20/2011).
- Lucia, S., Herrmann, L., & Killias, M. (2007). How important are interview methods and questionnaire designs in research on self-reported juvenile delinquency? An experimental comparison of Internet vs paper-and-pencil questionnaires and different definitions of the reference period. *Journal of Experimental Criminology*, *3*, 39–64.
- Thornberry, T. P., & Krohn, M. D. (2000). The self-report method for measuring delinquency and crime. In D. Duffee (Ed.), *Measurement and analysis of crime and justice* (Criminal Justice 2000, Vol. 4, pp. 33–83). Washington, DC: National Institute of Justice, Office of Justice Programs, US Department of Justice.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, *60*(2), 275–304.
- Trapl, E. S. (2007). *Understanding adolescent survey responses: Impact of mode and other characteristics on data outcomes and quality*. Cleveland: Case Western Reserve University.
- Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, *280*, 867–873.
- Vereecken, C. A., & Maes, L. (2006). Comparison of a computer-administered and paper-and-pencil-administered questionnaire on health and lifestyle behaviors. *Journal of Adolescent Health*, *38*(4), 426–432.

- Weisband, S., & Kiesler, S. (1996). Self-disclosure on computer forms: Meta-analysis and implications. In R. Bilger, S. Guest, & M. J. Tauber (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems: Common ground* (pp. 3–10). Vancouver: ACM.
- Wright, D. L., Aquilino, W. S., & Supple, A. J. (1998). A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *Public Opinion Quarterly*, 62(3), 331–353.

Simone Walser holds a Master in psychology. She is currently senior researcher and Ph.D. student at the University of Zurich, Institute of Criminology. Her research interests include juvenile delinquency and particularly the effects of situational factors and routine activities on youth behavior.

Martin Killias is professor of criminal law and criminology at the University of Zurich. His work centers on international crime and self-report surveys and evaluations including RCTs. He is a member of the Steering Committee of the Campbell Collaboration Crime and Justice Group.