Transportation Research Part F: Traffic Psychology and Behaviour,Volume 13, Issue 2, March 2010, Pages 106-114

The effect of social desirability on self reported and recorded road traffic accidents

A.E. af Wåhlberg* (a)
L. Dorn (b)
and
T. Kline (c)

a. Department of Psychology
Uppsala University
Uppsala, Sweden

b. Department of Systems Engineering and Human Factors,
Cranfield University
Cranfield, United Kingdom

c. University of Calgary
Calgary, Canada

*** Corresponding author**

Abstract

The use of lie scales has a fairly long history in psychometrics, with the intention of identifying and correcting for socially desirable answers. This is one type of common method variance (when both predictors and predicted variables are gathered from the same source), which may lead to spurious associations in self-reports. Within traffic safety research, where self-report methods are used abundantly, it is uncommon to control for social desirability artifacts, or reporting associations between lie scales, crashes and driver behaviour scales. In the present study, it was shown that self-reports of traffic accidents were negatively associated with a lie scale for driving, while recorded ones were not, as could be expected if the scale was valid and a self-report bias existed. We conclude that whenever self-reported crashes are used as an outcome variable and predicted by other self-reports, a lie scale should be included and used for correcting the associations. However, the only lie scale now in existence for traffic safety is not likely to catch all socially desirable responding, because traffic safety may not be desirable for all demographic groups. New lie scales should be developed specifically for driver questionnaires, to counter this suspected bias and artifactual results.

Key words: common method variance, social desirability, DSDS, methodology, self-report, accident

1. Introduction

Methodology using self-report data in psychology has been popular for a long time in most sub-areas of the discipline. The advantages of such an approach include the ability to ask any type of question, economy, and (apparent) simplicity of use. However, it has also been recognized that self-reports may be susceptible to what is called common method variance, i.e. biases of the (single) data source which influence several variables and create, increase or decrease associations between variables. If such biases exist, artefactual effects will result.

Social desirability (i.e. a tendency to report in a way that make the respondent look good) (Edwards, 1957; Nunnally & Bernstein, 1994) is one instance of a group of social/cognitive biases which can create common method variance in self-reports (Campbell & Fiske, 1959). To counter this tendency, so-called lie scales were developed and introduced into psychometric tests (e.g. Crowne & Marlow, 1960). These were thought to be able to detect faked replies. This was accomplished by measuring replies to items with a fairly obvious socially desirable content, like 'I have never stolen a thing in my life, not even a hairpin'. People who endorsed such items could be suspected of not really telling the truth, with or without their own knowledge. Using this basic logic, several different scales for socially desirable responding have been developed and validated. However, the standard type of validation is to test whether differences are found between so-called 'Fake good' and 'Standard' conditions (e.g. Blake, Valdiserri, Neuendorf & Nemeth, 2006). Such a test does not yield any information about whether there are individual differences in lying, and whether this influences other reports. Despite this shortcoming of the validity testing, when the influence of a socially desirable responding on other individual differences variables is tested, the validity of the lie scale would often be taken for granted (e.g. Ferrando, 2008).

Effects of socially desirable responding have been found in several research areas, e.g. organizational research (Moorman & Podsakoff, 1992), goal orientation (Tan & Hall, 2005) and personality (Ferrando, 2008). However, it should be noted that the main problem is not socially desirable responding as such, but the use of a common method and source for gathering both dependent and independent variables. If data sources differ between these variables, the problem would seem to be small (Pauls & Stemmler, 2003; Kurtz, Tarquini & Iobst, 2008).

Within traffic psychology, the use of self-reports (for both independent and dependent variables) has also been popular, and would seem to become ever more so. Here, the risk of common method variance due to social desirability would seem to be great. Lie scales, however, have rarely been a part of this research (the exceptions include Williams, Henderson & Mills, 1974; Donovan, Queisser, Salzberg & Umlauf, 1985; Lajunen, Corry, Summala & Hartley, 1998; Dula & Ballard, 2003; Fernandes, Job & Hatfield, 2007; Dorn & Gandolfi, 2008; Wickens, Toplak & Wiesenthal 2008), even though many of the attitudes and behaviours studied in traffic safety research have very clear social implications. However, although several authors have acknowledged the possible influence of social desirability on self-reports of behaviours and attitudes, few would seem to have thought that it would affect reports of crashes and other dependent variables (for a review and discussion, see af Wåhlberg, in press), which is necessary.for CMV effects.

The only lie scale developed especially for driver behavior would seem to be the Driver Social Desirability Scale (DSDS), which was constructed by Lajunen, Corry, Summala and Hartley (1997), and contains two sub-scales, Driver Impression Management (DIM) and Driver Self-Deception (DSD). The first focuses upon faking, the other on overly positive beliefs about oneself. The items are similar in content to those used in other lie scales, with rather improbable descriptions of behaviour, or lack thereof (see the items in the Appendix). However, although it is, of course, a positive development that some recognition of the social desirable responding phenomenon has belatedly come to traffic psychology, the DSDS has received little attention (the exceptions include Lajunen & Summala, 2003; Caird & Kline, 2004; Sundström, 2008; Dorn & Gandolfi, 2008). Given the prodigious number of studies published within traffic safety each year based solely on self-reported data, a validated lie scale for this type of reporting would seem to be needed.

However, validating a lie scale is not simple, especially if the standard method of 'fake good' is not accepted as valid. If the behaviour of interest is not expected to correlate with social desirability, but does have some sort of social implication, a lie scale can be tested against self-reports of this behaviour, and any association would constitute evidence in favour of the validity of the scale.

However, for the DSDS, the situation is somewhat more complicated, because the items used in it are actually desirable from a safety point of view (see Apppendix). Drivers who claim to behave in these ways may therefore be responding honestly, and the low numbers of crashes they also report are a logical consequence of these safe behaviours. In the end, a negative correlation between the DSDS and self-reported crashes does not constitute evidence in favour of its validity and/or a bias in accident reporting.

Lajunen et al. (1997) reported correlations of about .20 between the DSDS scales and self-reported accidents where the driver was to blame. For not responsible accidents, the effects were much weaker. In another study by the same authors (1998), the effect of the DIM scale in a hierarchical multiple regression was not significant. This might be seen as an indication of a bias in reporting, but need further corroboration.

The method suggested is to test recorded as well as self-reported crashes against the DSDS scale. If self-reported accidents correlates negatively with the DSDS scale, but recorded ones do not (or positively), the hypothesis that social desirability contaminates self-reported collision data would be supported, and the DSDS scale validated, because such a result would rule out the competing hypothesis that drivers who respond on the socially desirable end of the scale are in reality better drivers.

In summary, this study was intended to investigate the association of socially desirable responding, as measured by the DSDS, with self-reported and recorded number of traffic accidents, with the aim of validating the driver lie scale and investigating whether there exist a systematic bias in self-reports of road traffic accidents. If corroborated, this hypothesis would mean that CMV effects might exist in all studies that have used self-reports to predict self-reported crashes.

2. Method

2.1 General

Three sources of data were available for the present study, covering many different driver groups and two countries. DriverMetrics at Cranfield University (UK) has developed several online driver risk assessment instruments for driver education purposes (Dorn & Gandolfi, 2008; Dorn & Garwood, 2005) and all assessments include the DSDS scale. Also from the UK, and gathered online, was one sample of young offending drivers and one random sample. Both these were part of an evaluation of a Driver improvement course for young traffic offenders, developed by the e-training company a2om and delivered by the driver trainers DriveTech. Finally, data from Canada, used in a previous study (Caird & Kline, 2004) was included. These samples are demographically described in Table 1.

2.2 Samples

2.2.1 DriverMetrics samples

The DriverMetrics data covered four bus-driver samples from a major company, which had been gathered for the development of a bus driver profiling system and a bus driving simulator study. All participants were operational bus drivers with at least 3 months experience. Participants were asked to complete the questionnaire online prior to taking part in the bus driving simulator study. There were four samples of fleet drivers, and two police driver groups. All these samples were gathered in the UK The sample called Bus drivers 3 consisted of newly recruited chauffeurs, and they reported upon all their driving mishaps.

For the police drivers, the items were somewhat modified to exclude emergency conditions.

2.2.2 Young and Private driver samples

The evaluation of the UK driver improvement program, utilizing on-line teaching, included a questionnaire as part of the evaluation process. The first sample were young traffic offenders (mainly speeding offences), and were offered the training as an alternative to a fine and/or losing their license. The second sample was a random control group who were recruited using an e-marketing scheme. Both these samples were approached a second time, and so two measurements were available for a sub-group of each. The young drivers were all active drivers, as they had been caught offending in traffic. In the random sample, half of the drivers reported the lowest category (200 miles/month), which in principle mean that they could be non-drivers. The young drivers were asked to respond to the first wave of the questionnaire before starting the course, and could not log in to their first online teaching module without having done this. After having finished the last module, they were automatically linked to the second wave of the questionnaire, but this version did not contain the DIM scale or an item about accidents, and so these data were not used here. For the third wave responses, an e-mail was sent six months after the course, asking the drivers to respond again. About twenty percent of drivers did so (the numbers in the present tables are not indicative of the real response rate, as there is a time lag of about seven months between the waves, and the project was still running at the time of the present paper being written). All respondents were informed that their responses would only be used for research. Further information about this project and samples can be found in af Wåhlberg (submitted, a).

2.2.3 Truck driver sample

Finally, one sample of Canadian truck drivers was included.
These drivers were all active drivers for the truck company that supplied the recorded information.

Table 1 about here

2.3 Measures

The original DSDS had twelve items (see Appendix). In the present data, not all of these were used in all samples. Instead, there was a range from four to twelve items used, and in one instance the data did not include item-level information, but only total values for the DIM scale (see Table 2).

2.4 Accident data
All self-reported crash data was for all kinds of collisions, regardless of culpability, type of vehicle and private or professional driving. The time periods reported for are stated in Table 1. Within the bus and truck companies, their own definitions of mishaps and fault were used.

Table 2 about here

2.5 Statistical methods
As the DIM and DSD scales were fairly normally distributed (see Table 2), while the crash variables ranged from dichotomized with very skewed distributions to half-way normal (for the bus drivers), three different methods were used for all samples; correlations, t-tests and effect size (d) calculations. For the latter two, the grouping was by the mean of the lie scale. Correlations and t-tests are fairly common within the accident prediction literature, and the effects reported here are, therefore, comparable to older studies. Reports from the last decade, on the other hand, tend to use advanced statistical modelling, without reporting zero-order effects, and are therefore not quantatively comparable to anything. Also, different statistical methods tend to yield different results, and are preferred by different researchers, wherefore it was deemed necessary to do a more comprehensive analysis than is usually the case.
Pearson correlations are often deemed not fitting for skewed data. However, in previous work, no distortions of the size of effects have been found, even when the distributions have been extremely skewed. Instead, the size of correlations mostly conform to a linear function derived from a hypothesis which is unrelated to the statistical assumptions and characteristics of the Pearson correlation (af Wåhlberg & Dorn, 2007; 2009; af Wåhlberg, in press).

3. Results
For each sample with item data, the item values were summed to form the DSD, DIM and DSDS scales. Cronbach's alpha values were computed for all of these (see Table 2), and most were found to be acceptable. For one sample of bus drivers, the item-level data was not available, but only the overall DIM score.
The associations between age, experience and the DSDS scale were computed, as well as the correlations between the DIM and DSD sub-scales (Table 2). It can be seen that the sub-scales were moderately associated in all samples, while the results for age/experience were mixed with some samples showing positive and others negative relationships. Thus, we have no evidence of a consistent pattern of responding to the social desirability scales for these particular demographic variables.
Next, the associations between social desirability and accidents (recorded and self-report) were calculated, using Pearson correlation, t-tests and effect sizes (after dichotomization of the crash variable). The results can be seen in Tables 3-4. Finally, the mean effects over all samples for all tests were computed, separately for recorded

and self-reported data, excluding the crash-free kms variable. It can be seen that the association for recorded data was very close to zero, but on the positive side, while the self-reports yielded a negative effect across all samples.
A similar conclusion can be reached if only significant values are studied. All these were in the expected directions.

Tables 3-4 about here

4. Discussion

The results of the present study support the notion that a reporting bias for self-reports of accidents may exist, which can be identified by the DSDS. Perhaps the most telling effect is that of mainly negative correlations for the self-reported accidents, while those with the recorded crash variables tended to be positive, although very close to zero. Differences in effects between recorded and self-reported accident data are usually claimed to be due to the lesser variance of the first. Such an explanation is not viable here, as there is a difference in direction for the effects. Therefore, if the variance in recorded data was larger, the difference would increase.
These findings add to the small, but important literature on social desirability in traffic accident research. Donovan, Queisser, Salzberg and Umlauf (1985) did not find any differences in social desirability between groups contrasted by accidents from state records, and Williams, Henderson and Mills (1974) also reported no effect for traffic offenders (data from records). This lack of effect could be expected from the present hypothesis. The results reported here are therefore not entirely new, but the present study would seem to be the first to test for differences between recorded and self-reported data.
It could, of course, be argued that the associations found are weak, and thus not very important. However, most correlations with accidents that have been reported in the literature are also weak, and would, if not be totally eradicated, at least be significantly attenuated if social desirability effects were held constant. For example, in the meta-analysis of the associations between the Big Five personality constructs and accidents by Clarke and Robertson (2005), almost half the studies used self-reported accidents as the outcome variable. The raw mean correlations ranged from .098 to .182, and a correction for social desirability might therefore have very significant effects upon this association. Also, in a separate paper utlizing the Young and Private drivers samples used here, it has been shown that the DIM scale explain half of the variance shared between self-reported crashes and several well-known driver inventories (af Wåhlberg, submitted, b). Similar results were reported for the Crowne-Marlowe lie scale by Willemsen, Dula, Declerq and Verhaege (2008).
In many accident prediction studies utilizing questionnaires, the authors state that social desirability or similar concepts might have influenced the independent variables (e.g. Brown & Cotton, 2003), although no direct assessment of this effect was included. However, acknowledgement of this source of possible bias has seldom been extended to the dependent variable (exceptions include Pöysti, Rajalin & Summala, 2005), and when it has, it has usually been claimed that socially desirable responding should lead to under-estimations of the real associations between accidents and predictors (West, 1995; Hatakka, Keskinen, Katila & Laapotti, 1997; Lawton, Parker, Stradling & Manstead, 1997; Sullman, Meadows & Pajo, 2002), due to limited variation. That shared variance is created by this type of bias in reporting does not seem to have been understood, as few attempts to include some sort of lie scale and

holding the effect constant has been found (among the few are Arthur & Graziano, 1996). In some studies, it is argued that social desirability should not have been a problem, due to the anonymity of the respondents (Lajunen & Parker, 2001). However, whether this method really has the desired effect would currently seem to be uncertain, and research on this needed.

It is very common within traffic safety research not to use accidents as the outcome variable, ostensibly due to its problematic statistical properties (infrequent occurrence). Instead, various ‘risky driving’ measures are used as outcome measures, because it is assumed that these risky behaviours are directly related to accidents (an assumption which is unproven, at best, af Wåhlberg, in press). This approach however, does not erase the social desirability effects described. If anything the problem is likely to be even worse using such measures. While accidents are uncommon and easily counted (although not necessarily remembered or truthfully reported upon), speeding and tailgating, for example, are rather difficult to recall precisely, because they appear in much more nebulous categories such as ‘often’, ‘seldom’ etc. It is easy for a respondent to change his/her reply a bit without feeling dishonest, because the discrepancy is not as evident to oneself as if a remembered crash was not reported (or an extra one added). It should also be remembered that many ‘risky driving behaviour’ scales that are used as outcome variables started out as crash predictors (notably the DBQ), and that these are very prone to social desirability effects, which that their intercorrelations to a large degree are due to reporting bias (af Wåhlberg, submitted, b).

The present study did find that the DSDS does work as expected. However, despite this outcome, the problem of the item content remain to some degree. As can be seen in the Appendix, most DSDS items (especially the DIM scale) are socially desirable behaviours from the traffic safety specialist’s point of view. That such behaviours are seen as the social norm by researchers can be exemplified by Lajunen and Summala (2003), who said about the Manchester Driver Behaviour Questionnaire; “...driving behaviours listed in the DBQ are all socially non-desirable behaviours...” (p. 104). This view is very much in line with the traditional way of constructing lie scales, where the overall social norm is used as the foundation for the construct. However, it can be questioned whether this norm is accepted by all drivers, or if there are subgroups that will not endorse such behaviours as desirable. However, this line of reasoning does not mean that the self-reports of such non-normative drivers are free from bias, only that it cannot be detected by using a traditionally constructed lie scale. What is suspected is that there could exist over-reporting of crashes and behaviours by sub-groups of drivers with very different attitudes towards driving than what is usually assumed to be the standard.

There could also exist another problem with the DSDS, due to its item content. It could actually be the case that there exist a small sub-group who honestly can say that they have never done the behaviours described in the items, and who also have rather few accidents.

Furthermore, the DSDS does not necessarily capture social norms within all driver groups, and effects should not be expected in all populations. It could be expected that within driver groups where crashes are seen as an inevitable part of driving, socially desirable responding as measured by the DSDS would not exist.

Given these suspected biases, the construction of different types of driver social desirability scales is recommended, as well as other validity control methods.

No consistent differences in effects were found between the subscales, or the full version versus a subset of the items. From these results, there does not seem to be an indication that both scales or all items are needed.
Some researchers would argue that records of accidents are so unreliable (and often hard to access) as to be useless, and that self-reports are therefore a better option. However, such an argument rests upon the errenous conclusion that the proven under-reporting into, and known biases of, (state) records somehow make self-reports more valid. In fact, there exist very few comparative studies on this subject, and what little evidence there is indicates that the correlations between sources are suspiciously low, and that they yield different results when used as outcomes (af Wåhlberg, in press; af Wåhlberg, Dorn & Kline, in press).
Given the large number studies that have used only self-reported data in traffic psychology, the effects of socially desirable responding could be dramatic on the overall interpretation of results. Until further research has established more firmly how strong the effects described are, it is recommended that
a) self-reported accidents are not used for research, unless the predictors are gathered from another source,
b) if the use of self-reported collisions cannot be avoided, it is recommended that researchers include a lie scale, and partial out this effect, and
c) proxy dependent variables, like speeding, should also be controlled for social desirability effects.
Finally, it should be remembered that social desirability is just one of a group of common method variance effects that can create artifactual associations in self-reported traffic safety data. There is an urgent need to develop and implement methods to detect and counteract such threats to validity in the majority of traffic safety studies published. The alternative is that traffic safety research is built on a foundation of knowledge stemming from potentially biased and distorted data. .

Acknowledgement
The data on young offending drivers was gathered as part of the Thames Valley Police Young Driver Scheme evaluation, and the analyses undertaken here was part of the general research plan.

5. References

Arthur, W., Jr., & Graziano, W. G. (1996). The Five-factor model, conscientiousness, and driving accident involvement. *Journal of Personality, 64,* 593-618.

Blake, B. F., Valdiserri, J., Neuendorf, K. A., & Nemeth, J. (2006). Validity of the SDS-17 measure of social desirability in the American context. *Personality and Individual Differences, 40,* 1625-1636.

Brown, S. L., & Cotton, A. (2003). Risk-mitigating beliefs, risk estimates, and self-reported speeding in a sample of Australian drivers. *Journal of Safety Research*, *34,* 183-188.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Caird, J. K., & Kline, T. J. (2004). The relationships between organizational and individual variables to on-the-job driver accidents and accident-free kilometres. *Ergonomics, 47,* 1598-1613.

Clarke, S., & Robertson, I. T. (2005). A meta-analytic review of the Big Five personality factors and accident involvement in occupational and non-occupational settings. *Journal of Occupational and Organizational Psychology, 78,* 355–376.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349-354.

Donovan, D. M., Queisser, H. R., Salzberg, P. M., & Umlauf, R. L. (1985). Intoxicated and bad drivers: Subgroups within the same population of high-risk drivers. *Journal of Studies on Alcohol, 46,* 375-382.

Dula, C. S., & Ballard, M. E. (2003). Development and evaluation of a measure of dangerous, aggressive, negative emotional, and risky driving. *Journal of Applied Social Psychology, 33*, 263-282.

Dorn, L, & Gandolfi, J. (2008) Designing a Psychometrically-Based Self Assessment to Address Fleet Driver Risk. In Dorn L (Ed) *Driver Behaviour and Training Vol III.* Human Factors in Road and Rail Safety. Ashgate: Aldershot.

Dorn, L, & Garwood, L. (2005). Development of a psychometric measure of bus driver behaviour. *Behavioural Research in Road Safety: 14th Seminar.* Department for Transport: London.

Edwards, A. L. (1957). *The Social Desirability Variable in Personality Assessment and Research.* New York: Dryden.

Fernandes, R., Job, R. F., & Hatfield, J. (2007). A challenge to the assumed generalizability of prediction and countermeasure for risky driving: Different factors predict different risky driving behaviors. *Journal of Safety Research, 38,* 59-70.

Ferrando, P. J. (2008). The impact of social desirability bias on the EPQ-R item scores: An item response theory analysis. *Personality and Individual Differences, 44,* 1784-1794.

Hatakka, M., Keskinen, E., Katila, A., & Laapotti, S. (1997). Self-reported driving habits are valid predictors of violations and accidents. In T. Rothengatter & E. C. Vaya, *Traffic and Transport Psychology, Theory and Application*, pp. 295-303. Amsterdam: Pergamon.

Kurtz, J. E., Tarquini, S. J., & Iobst, E. A. (2008). Socially desirable responding in personality assessment: Still more substance than style. *Personality and Individual Differences, 45,* 22-27.

Lajunen, T., Corry, A., Summala, H., & Hartley, L. (1997). Impression management and self-deception in traffic behaviour inventories. *Personality and Individual Differences*, *22*, 341-353.

Lajunen, T., Corry, A., Summala, H., & Hartley, L. (1998). Cross-cultural differences in drivers' self-assessments of their perceptual-motor and safety skills: Australians and Finns. *Personality and Individual Differences*, *24,* 539-550.

Lajunen, T., & Parker, D. (2001). Are aggressive people aggressive drivers? A study of the relationship between self-reported general aggressiveness, driver anger and aggressive driving. *Accident Analysis and Prevention, 33,* 243-255.

Lajunen, T., & Summala, H. (2003). Can we trust self-reports of driving? Effects of impression management on driver behaviour questionnaire responses. *Transportation Research Part F, 6,* 97-107.

Lawton, R., Parker, D., Stradling, S. G., & Manstead, A. S. (1997). Predicting road traffic accidents: The role of social deviance and violations. *British Journal of Psychology, 88,* 249-262.

Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organizational Psychology, 65,* 131-149.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.

Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences, 35,* 263-275.

Pöysti, L., Rajalin, S., & Summala, H. (2005). Factors influencing the use of cellular (mobile) phone during driving and hazards while using it. *Accident Analysis and Prevention, 37,* 47-51.

Sullman, M. J., Meadows, M. L., & Pajo, K. B. (2002). Aberrant driving behaviours amongst New Zealand truck drivers. *Transportation Research Part F, 5,* 217-232.

Sundström, A. (2008). Self-assessment of driving skill - A review from a measurement perspective. *Transportation Research Part F, 11,* 1-9.

Tan, J. A., & Hall, R. J. (2005). The effects of social desirability bias on applied measures of goal orientation. *Personality and Individual Differences, 38,* 1891-1902.

West, R. J. (1995). Individual differences in accident risk: a review of findings and an examination of methods. In *Proceedings of Behavioural Research in Road Safety VI.* London, England, 4–5 September 1995.

Wickens, C. M., Toplak, M. E., & Wiesenthal, D. L. (2008). Cognitive failures as predictors of driving errors, lapses, and violations. *Accident Analysis and Prevention, 40,* 1223-1233.

Willemsen, J., Dula, C. S., Declerq, F., & Verhaeghe, P. (2008). The Dula Dangerous Driving Index: An investigation of reliability and validity across cultures. *Accident Analysis and Prevention, 40,* 798-806.

Williams, C. L., Henderson, A. S., & Mills, J. M. (1974). An epidemiological study of serious traffic offenders. *Social Psychiatry, 9,* 99-109.

af Wåhlberg, A. E. (in press). *Driver Behaviour and Accident Research Methodology; Unresolved Problems.* Aldershot: Ashgate.

af Wåhlberg, A. E. (submitted, a). Re-education of young driving offenders; effects on self-reports of driver behaviour.

af Wåhlberg, A. E. (submitted, b). Social desirability effects in driver behaviour inventories.

af Wåhlberg, A. E., & Dorn, L. (2007). Culpable versus non-culpable traffic accidents; what is wrong with this picture? *Journal of Safety Research, 38,* 453-459.

af Wåhlberg, A. E., & Dorn, L. (2009). Bus driver accident record; the return of accident proneness. *Theoretical Issues in Ergonomics Science, 10,* 77-91.

af Wåhlberg, A. E., Dorn, L., & Kline, T. (in press). The Manchester Driver Behaviour Questionnaire as predictor of road traffic accidents. *Theoretical Issues in Ergonomics Science,*

Appendix: The Driver Social Desirability Scale (Lajunen, Corry, Summala & Hartley, 1997)

Originally seven point scale, from 1: Not true, to 7: Very true

Driver Impression Management scale

1. I have never exceeded the speed limit
2. I have never wanted to drive very fast
3. I have never driven through a traffic light when it has just been turning red
4. I always obey traffic rules, even if I am unlikely to get caught
5. I always keep sufficient distance from the car in front of my car
6. If there were no police controls, I would still obey speed limits
7. I have never exceeded speed limits or crossed a solid white line in the centre of the road when overtaking

Driver Self-Deception Scale

8. I always know what to do in traffic situations
9. I never regret my decisions in traffic
10. I don't care what other drivers think of me
11. I always am sure how to act in traffic situations
12. I always remain calm and rational in traffic

Table 1: Descriptive data for the samples included in the study (mean/std and percent). Country of origin, number of subjects, mean/std of age, experience, number of accidents (differing time periods), and percent women in the sample.

Subjects	N	Age	Experience (years)	Mean number of accidents (all)	Time period for accidents	Mean number of accident per year (all)	Accident source	Sex
New bus drivers	142	37.9/10.0	16.4/9.5 #	1.40/1.62	9 months	1.87/2.16	Record	12%
Bus drivers 1	95	48.7/10.0	12.9/11.8 (years of employment)	3.14/2.76	Variable	0.80/1.13 (N=91)	Record	6%
Bus drivers 2	293	46.1/10.4	12.6/10.9 (PCV license)	2.96/2.73	Variable	0.84/1.26	Record	7%
Bus drivers 3	51	-	13.5/11.1 #	0.10/0.30	XXX3 years	XXX	Self-report	22%
Police drivers 1	313	36.3/8.7	20.3/8.7 (N=306)	0.46/0.63 (dichotomized scale)	3 years	0.15/0.21	Self-report	16%
Police drivers 2	381	a	b	0.23/0.56	3 years	0.08/0.19	Self-report	16%
Fleet drivers 1	1795	40.7/9.5	18.2/9.0 #	0.34/0.70	3 years	0.11/0.23	Self-report	18%
Fleet drivers 2	523	40.9/10.2	17.0/9.9	0.27/0.60	3 years	0.09/0.20	Self-report	22%
Fleet drivers 3	803	42.6/9.7	19.7/8.9 #	0.25/0.56	3 years	0.08/0.19	Self-report	20%
Fleet drivers 4	168	37.9/9.2	18.9/9.0 (N=158) c	0.57/0.95	3 years	0.19/0.31	Self-report	55%
Young drivers Second wave	4453	21.7/2.2	3.2/2.1	0.58/0.87	Variable Six months	0.27/0.71	Self-report	40%
Private drivers Second wave	1231	33.2/14.1	12.7/13.0 #	0.84/1.36	Variable Six months	0.12/0.86	Self-report	44%
Truck drivers	111	42.6/6.7	26.8/6.9 (XXX)	0.08/0.33 0.57/0.71	Two years Variable	0.04/0.17 0.02/0.03	Self-report Record	4%

a Age data was only available in four broad categories (20-35, 36-41, 42-47, 48+ years). Fairly similar numbers of the sample belonged to each (28, 25, 24, 23 percent)
b As with age for this sample, experience was recorded in broad categories; <15, 16-25, 26-30, 30+ years. Thirty-nine percent were in the second category.
PCV=Personal Commercial Vehicle
c This variable had several missing values. One was deleted, as it was equal to the driver's age.
years of driver's license

Table 2: The means, standard deviations, kurtosis and alpha values of the scales in all samples. Also, the correlations between age, experience and the DSDS scale, and the intercorrelations between the subscales. In samples Young drivers and Private drivers, the second row are for re-testing of those in the first wave.

Subjects	N	DSDS scale	Steps of scale	Mean/std	Kurtosis	Skewness	Alpha	Correlation with age	N for experience	Correlation with experience	Correlation DIM/DSD
New bus drivers	142	DIM	1-7	(28.6/9.82)	-0.15	-0.01	-	.016	142	-	-
Bus drivers 1	95	DSDS DIM (7 items) DSD (5 items)	1-7	4.17/1.00 3.92/1.25 4.52/1.05	0.29 -0.50 0.76	-0.19 -0.15 -0.46	.83 .82 .71	.113	93	.051 (years of employment)	.444***
Bus drivers 2	293	DSDS DIM (7 items) DSD (5 items)	0-1	0.50/0.24 0.44/0.26 0.59/0.31	-0.59 -0.54 -0.85	-0.12 0.15 -0.46	.75 .68 .64	.057 .071 .022	287	.027 -.015 .068 (years of PCV license)	.412***
Bus drivers 3	51	DSDS DIM (7 items) DSD (5 items)	1-6	3.67/1.07 3.52/1.25 3.87/1.37	-0.05 -0.61 0.56	-0.62 -0.41 -0.81	.82 .79 .81	-	51 #	-.240 -.21 -.18	.346*
Police drivers 1	316	DIM (7 items)	5-1 (reversed)	3.61/0.98	1.19	0.92	.81	-.056	306 #	.070	-
Police drivers 2	381	DSDS (8 items)	1-5	2.23/0.55	-0.06	0.30	.68	-.150	-	-	-
Fleet drivers 1	1795	DSDS DIM (6 items) DSD (4 items)	0-6	3.23/1.09 2.71/1.35 4.00/1.18	-0.33 -0.58 0.18	0.08 0.19 -0.54	.83 .81 .82	.009	1795 #	-.020	.407***
Fleet drivers 2	523	DSDS DIM (6 items) DSD (4 items)	0-6	3.07/1.10 2.64/1.36 3.71/1.18	-0.16 -0.45 0.18	0.20 0.26 -0.43	.84 .83 .81	-.042 -.069 .021	523 #	-.016 -.083 .105*	.425***
Fleet drivers 3	803	DSDS DIM (6 items) DSD (4 items)	0-6	3.35/1.06 2.94/1.30 3.96/0.33	-0.22 -0.48 0.33	-0.17 -0.03 -0.59	.81 .77 .82	.094** .063 .111**	803 #	.088* .042 .132***	.425***
Fleet drivers 4	168	DSDS DIM (7 items) DSD (5 items)	1-0 (reversed)	0.67/0.20 0.73/0.20 0.59/0.30	-0.58 -0.44 -1.36	-0.18 -0.47 -0.05	.68 .51 .67	-.009 -.101 .079	158 #	-.030 -.092 .038	.324***
Young drivers	4453	DIM (7 items)	1-5	3.14/0.91	-0.73	-0.07	0.82	-.059	4453 #	-.123***	-
Private drivers	1461 239	DIM (7 items)	1-5	2.95/1.03 2.87/1.00	-0.89 -0.79	0.22 0.42	.87 .86	-.090*** (N=1449) -.124	1232 # 203	-.080 -.123	-

								(N=233)			
Truck drivers	111	DSD (4 items)	1-5	2.87/0.54	-0.67	0.13	.52	-.359***	111	-.353***	-

* p<.05, p<.01, *** p<.001
# years of driver’s license
PCV=Personal Commercial Vehicle

Table 3: Results for samples where accident data was recorded by an organisation. Shown are type of drivers, DSDS scale(s) used and alpha values of these (if available), type of accidents, and N of sample. Thereafter, association between social desirability and accidents, calculated as Pearson correlation, independent t-test and Cohen's d. t-values and d computed for groups dichotomized by the mean on the DSDS variable. Positive values mean a negative association for these tests, apart from the crash free kilometers variable. Effect sizes (d) calculated with the low-DSDS group standard deviation as the denominator.

Subjects	N	DSDS variables	Accident type	Correlation	t	d
New bus drivers (9 months of driving)	142	DIM	At fault	.138	-0.06	-0.01
			All	.009	1.24	0.20
Bus drivers 1	95	DSDS	At fault	-.045	-0.28	-0.05
		DIM (7 items)		-.048	-0.21	-0.04
		DSD (5 items)		-.023	-1.43	-0.30
	91	DSDS	At fault/year	.095	-1.18	-0.39
		DIM (7 items)		.077	-0.96	-0.31
		DSD (5 items)		.088	-1.58	-0.54
Bus drivers 1	95	DSDS	All	-.130	0.28	0.05
		DIM (7 items)		-.137	0.22	0.04
		DSD (5 items)		-.068	-0.54	-0.11
	91	DSDS	All/year	.060	-0.96	-0.26
		DIM (7 items)		.073	-0.90	-0.25
		DSD (5 items)		.034	-1.38	-0.39
Bus drivers 2	293	DSDS	At fault	.012	-2.19*	-0.34
		DIM (7 items)		.006	-1.55	-0.23
		DSD (5 items)		.016	-0.85	-0.12
		DSDS	At fault/year	-.010	-1.48	-0.21
		DIM (7 items)		.024	-1.16	-0.16
		DSD (5 items)		-.046	0.54	0.06
Bus drivers 2	293	DSDS	All	-.067	-0.79	-0.09
		DIM (7 items)		-.085	-0.40	-0.05
		DSD (5 items)		-.023	-0.36	-0.04
		DSDS	All/year	-.080	-0.75	-0.09
		DIM (7 items)		-.059	-0.42	-0.05
		DSD (5 items)		-.077	0.85	0.09
Truck drivers	111	DSD (4 items)	All	.082	-1.21	-0.26
			All/year	.137	-1.72	-0.47
			Crash free kms (all)	-.085	0.49	0.09

Mean (except crash free kms)				.010	-0.69	-0.15

* p<.05, ** p<.01, *** p<.001

Table 4: Results for samples where accident data was self-reported. Shown are type of drivers, DSDS scale(s) used and alpha values of these (if available), type of accidents, and N of sample. Thereafter, association between social desirability and accidents, calculated as Pearson correlation, independent t-test and Cohen's d. t-values and d computed for groups dichotomized by the mean on the DSDS variable. Positive values mean a negative association for these tests, apart from the crash free kilometers variable. Effect sizes (d) calculated with the low-DSDS group standard deviation as the denominator. The scales of samples Police drivers 1 and Fleet driver 4 were reversed to the same type of scaling as the other samples' (high SD=high value).

Subjects	N	DSDS variables	Accident type	Correlation	t	d
Bus drivers 3	51	DSDS	All	-.218	2.50*	0.49
		DIM (7 items)		-.108	0.51	0.13
		DSD (5 items)		-.270	1.46	0.32
Police drivers 1	316	DIM (7 items)	All	-.021	1.52	0.16
Police drivers 2	384	DSDS (8 items)	All	.005	1.05	0.12
Fleet drivers 1	1795	DSDS	All	-.048*	0.55	0.03
		DIM (6 items)		-.039	0.96	0.04
		DSD (4 items)		-.043	1.90	0.08
Fleet drivers 2	523	DSDS	All	-.050	-0.39	0.04
		DIM (6 items)		-.016	0.64	-0.06
		DSD (4 items)		-.089*	1.23	0.10
Fleet drivers 3	803	DSDS	All	-.091*	1.84	0.13
		DIM (6 items)		-.093**	1.46	0.10
		DSD (4 items)		-.053	0.33	0.02
Fleet drivers 4	167	DSDS	All	-.078	3.29**	0.45
		DIM (7 items)		-.097	1.75	0.23
		DSD (5 items)		-.033	0.60	0.10
Young drivers	4444	DIM (7 items)	All	-.129***	6.66***	0.19
			All/year	-.053***	3.57***	0.09
			All/mile	-.009	1.07	0.02
	342#		All/year	-.017	0.61	0.07
Private drivers	1230##	DIM (7 items)	All	-.177***	5.85***	0.33
			All/year	-.075**	1.49	0.10
			All/mile	-.056	1.29	0.09
	239#		All	-.112	1.77	0.18
			All/mile	-.115	1.89	0.18
Truck drivers	111	DSD (4 items)	All	.021	-0.44	-0.08
			Crash free kms (all)	.192*	-1.16	-0.22

Mean				-.100	1.67	0.14

* p<.05, ** p<.01, *** p<.001

repeated measurement of the larger sample.

One outlier on the collision per year variable deleted.