



Re-education of young driving offenders; Effects on self-reports of driver behavior

A.E. af Wåhlberg*

Department of Psychology, Uppsala University, P. O. Box 1225, 751 42 Uppsala, Sweden

ARTICLE INFO

Article history:

Received 10 June 2009

Received in revised form 19 October 2009

Accepted 4 May 2010

Available online 6 July 2010

Keywords:

Evaluation

Driver improvement

Questionnaire

Validity

Socially desirable responding

ABSTRACT

Introduction: Offending drivers are often re-educated, but these courses have seldom been shown to have any safety effects. **Method:** An on-line improvement course for offending drivers below the age of 25 was evaluated with several driver inventories. **Results:** The drivers reported higher levels of aggression, stress, sensation seeking, drunk driving, and driving violations, six months after the course than before. However, these levels were lower than those of controls, indicating that the initially low levels for the education group were due to socially desirable responding, as measured by a lie scale, an effect that waned after the course. **Discussion:** The results can be interpreted as a positive effect of the education, although this conclusion is tentative and not in agreement with all effects in the data. **Impact on industry:** The results are in disagreement with previous evaluation studies using the same or similar instruments, and show the need to include controls for social desirability in self-report studies.

© 2010 National Safety Council and Elsevier Ltd. All rights reserved.

1. Introduction

1.1. History

Traffic rules and laws were invented to increase safety of transport by governing the behavior of drivers, and there is thus a pronounced interest in making sure that road users actually adhere to these laws. However, although policing the roads does have an effect on driver behavior (Hakkert, Gitelman, Cohen, Doveh, & Umansky, 2001), there are few places within society where lawlessness seems to rule more widespread than among drivers. Apparently, not even the deterrents of possibly crashing, receiving a fine, penalty points, loss of license, and ultimately a jail sentence can keep some people from speeding, tailgating, driving drunk, and so forth. This creates a problem for the authorities. What should be done about drivers who are not impressed by the risks they run?

For many decades, driver training has played a large part in traffic safety work. Indeed, most drivers in industrialized countries seem to take at least some lessons from professional instructors before passing their driving test. However, the evidence in favor of

the practice of professional training¹ is very limited. For example, Ferdun, Peck, and Coppin (1967) found that those who had taken a driver training class had slightly more crashes than those who did not, in schools that offered such classes. Owsley, McGwin, Phillips, McNeal, and Stalvey (2004) reported similar results for safety education of older drivers.

Such results can be found in several reviews of evaluations of driver training and education; few studies have shown positive results, and some even report so-called perverse effects (i.e., increases in crash rates). In general, the research area of driver training features an amazing agreement among scientists, the conclusion in every review being that there is no evidence that any safety gains have been achieved by training or educating drivers (Christie, 2001; Klein, 1966; Lonero, 2008; Masten & Peck, 2004; Mayhew, Simpson, Williams, & Ferguson, 1998; Roberts, Kwan, & Cochrane Injuries Group Driver Education Reviewers, 2008). Very few researchers seem to have a different view (e.g., Stanton, Walker, Young, Kazi, & Salmon, 2007). Still, research into this problem is undertaken, and educational developments may change the present picture, as could studies utilizing larger samples and better methodology. As an alternative to

¹ Christie (2001) discussed the differences between training and education for drivers. The former mainly features technical skills, while the latter most often have no in-car modules at all, and instead is aimed at attitudes, knowledge, risk perception, and similar cognitive features. These differences may in practice, however, often be blurred, because a driver course can contain all these features, and the terms have often been used interchangeably. In the present paper, however, the distinction will be upheld, as far as possible.

* Tel.: +46 18 471 25 90, +46 18 33 90 95; fax: +46 18 471 21 23.

E-mail address: anders.af_wahlberg@psyk.uu.se.

URL: <http://www.psyk.uu.se/hemsidor/busdriver>.

crash rates, there is also the possibility of using behavior, attitudes, and other variables as outcome criteria. Such parameters are more sensitive to changes, but have the drawback of being weakly associated with crashes.

The concept of driver training/education has also been applied to drivers who commit driving offenses. The basic idea for these educational programs is the same as for fines and penalty points; offenses are thought to be indicators of elevated risk for crashing, and the fines and education are there to change the behavior of the driver. However, already at an early stage, the optimism regarding a kind of re-education of drivers were criticized as being unfounded (Johnson, 1939; Johnson & Cobb, 1938). The evaluations that showed beneficial results were simply misleading, mainly because they did not use control groups. It can be noted that, apparently, this early re-education attempt did not entail any practical teaching of driving skills, but was based upon changing the way the driver thought about his/her own driving.

As with basic driver training, the reviews regarding the effects of post-licensing training and education programs for offending drivers generally show no effect (Ker et al., 2005), in terms of reducing accident involvement. Some reviews have found reliable effects for violations but not for crashes (Janke, 1994; Kaestner, 1968; Lund & Williams, 1985). In Harano and Peck (1972) this type of result was described as a rule of driver improvement research. Perverse effects have, as stated, been found in some studies (see the review by Struckman-Johnson, Lund, Williams, & Osborne, 1989).

Only one review seems to have reported positive effects for crashes; McGuire and Kersch (1969) summarized the results of several studies in which it was found that education was more effective than the alternatives (and combinations of these), probation and fines. However, it should be noted that these effects were for fairly short periods (i.e., about a year) whereafter the effect disappeared. Here, it can be noted that the educational regiments used did not have a practical driving content, but mainly targeted perceptions and attitudes, through group or single discussions. Other studies reviewed by these authors did not find any effects for re-education of offenders. For various other road safety improvement schemes, results are equally disappointing (Carcallion, Rachid Salmi, & Atout-Route Evaluation Group, 2005; Nasvadi & Vavrik, 2007).

The lack of effects for driver improvement schemes may to some degree be due to the educational content being different between courses (i.e., there may be some sort of effective course content, but the lack of effects in non-effective ones have clouded this). Also, evaluation methods have differed, making the results difficult to compare. At times, the components have also been poorly described in the evaluation reports.

The majority of evaluations of driver improvement courses seem to have been undertaken in the United States, while there are also some fairly recent ones from Britain. It can be noted that the methodology tends to differ between these sources. However, before analyzing these differences, the methodology of evaluations of such courses must be discussed.

1.2. Evaluating driver improvement schemes

The basic question posed regarding all training and education is simple; does it have the intended effect? For driver improvement schemes, this means reducing collision rates. At times, researchers and traffic safety practitioners seem to imply that the goal is to change attitudes and/or behavior (which are here called proxy safety variables). Although this may in some sense be true, this is not the real goal, because such statements always carry with them the (often implicit) assumption that if we are able to change the proxy factors, we will also influence the crash rate of the drivers. It must be acknowledged, however, that this connection is an assumption, and that the evidence to date shows that there are no variables that are strongly associated with

risk of collision (af Wählberg, 2009), although the statistical properties of crashes does make it difficult to know exactly. This means that we might very well achieve changes in attitudes and behaviors, while the crash rate remains the same, or even increases.

This is also the case for driving offenses, a key variable in driver improvement. The correlation between offenses and crashes is usually .1–.2 (af Wählberg, 2009), meaning that less than 5% of the variance is shared. As a consequence, driving offenses are a very blunt instrument when it comes to identifying drivers at risk of crashes (between individual differences), but they are also a precarious measure for evaluation of driver improvement schemes (within-individual differences). As noted, a common effect is that a reduction of offenses is achieved, while collision rate remains the same (Lund & Williams, 1985), or even increases (Janke, 1994). A similar difference can be seen in Nasvadi (2007), as compared to Nasvadi and Vavrik (2007). In the first study, there were positive changes in self-reported driving behavior, while the other found no difference, or an increase, in recorded crashes.

Concerning variables to use as criteria in evaluations of driver training and education, collisions would therefore seem to be the most important one, while offenses and questionnaires (the main alternatives) can only be seen as very weak indicators of safety effects. Despite this fact, evaluations of driver training often use such outcome measures, and it is therefore important when reviewing previous studies to report exactly what kinds of measurement techniques and variables were used.

Also, it must be noted how the data have been gathered. With time, the use of self-reported traffic data has become very popular, both for behavior and crashes. However, it has repeatedly been shown that self-reports of crashes are unreliable, due to memory failures and various cognitive mechanisms that distort the responses (af Wählberg, 2009; af Wählberg, 2010; af Wählberg, Dorn, & Kline, 2010; af Wählberg, Dorn, & Kline, in press). For example, Planek, Schupack, and Fowler (1974) reported, regarding their evaluation of a defensive driving course; "With reference to accidents, a discrepancy appears to exist between the per cent of reduction in the self-reported data and the state records for study group respondents. State record data indicate a before-after reduction in accidents of approximately 15 per cent less than shown in self reports" (p. 295).

The main problem would seem to be socially desirable responding, where people answer questions in ways that they think will make them look good. It can be noted that being anonymous does not seem to eradicate this effect (af Wählberg, 2010; af Wählberg et al., 2010). The same kind of effect would seem to be present for drivers' responses to questions about their behavior on the road (af Wählberg, 2009), and common method variance effects follow (af Wählberg, 2010; i.e., associations are found in the data that are not due to real effects, but to various response biases). Such effects are well known in other research areas, like managerial and consumer psychology (for reviews, see Moorman & Podsakoff, 1992; Podsakoff, Mackenzie, Lee, & Podsakoff, 2003; Schmitt, 1994).

However, it should be pointed out here that the evidence concerning distorting mechanisms in self-reports, and the weak associations between various variables (e.g., offenses) and traffic safety mainly concern individual differences, while the situation is different for changes within individuals. As an example, claiming that driver A is more dangerous than driver B just because he/she drives 10 miles faster per hour is mainly wrong, because crash rate is influenced by many other variables than speed. On the other hand, if driver A would reduce his/her average speed by 10 miles an hour, this would most certainly make him/her a safer driver. The certainty in this prediction stems from the fact that, in principle, all other influences on this driver's safety remain the same.

In the end, it is therefore possible that other outcome measures than crashes can be used for evaluating driver improvement schemes, even self-reported ones. Such data might yield insights into behavior change

that may or may not transfer into increases in safety. However, great care must be put into avoiding artefactual results, something that is not simple, because there would seem to exist no research on how drivers react to the very special situation of being re-educated, most often due to having broken the law, and therefore being under a certain threat of prosecution. However, the most important issue is whether the drivers subjectively feel that they are in a situation where they can be identified and/or suffering adverse consequences. How people assess situations regarding these possibilities does not seem to be known.

Also, all proxy safety variables and self-report instruments seem to have been validated in individual differences research. To be shown to have validity as an outcome measure in a treatment study, a very different design would be needed, where the difference in accident rate for each individual was correlated with the change in offenses and other proxy variables. No such research has been located.

In summary, all proxy safety variables that have been used for evaluations have doubtful validity, for several different reasons. Recorded collisions, on the other hand, suffer from the well-known problem of under-reporting, as only a minor number of the crashes happening come to the attention of the police (the most common data source for evaluations). This leads to low statistical power of any study using such data, which must be countered by massive numbers in the groups tested. As a further problem here, it is probably not possible to extend the evaluation period to more than a year, as long-term effects hardly can be expected from re-education of drivers. It is therefore difficult to increase the variance in the sample by this method, and the only possible solution is to use large numbers of drivers.

In the present paper, only self-reported data are reported upon. Although widely used for research and evaluation purposes within traffic safety, self-reports cannot be regarded as trustworthy, and when calculations of individual differences are undertaken, the risk of common method variance effects is high (af Wählberg, 2009). Artefactual associations between driver inventory responses and self-reported collisions have been shown to exist within the presently used data (af Wählberg, 2010), and although within-individuals effects of this type in evaluations have not been reported, there seems to be a lack of research into this possible problem. The present research therefore had as its basic tenets that controls for social desirability were needed and that any results found would need further corroboration from other data sources.

The introductory review of evaluations of driver training and improvement schemes concerned (recorded) crashes only, with most results stemming from the United States, from the 1950s onwards. Turning to the alternative of self-reported, proxy safety variables, the studies using this method would seem to have the characteristics of being fairly new and British.

1.3. British evaluations of driver improvement schemes

As driver improvement schemes have been run in the United Kingdom for many years, some evaluations have been undertaken. In general, British evaluations seem to be more interested in proxy variables, like attitudes, as outcome measures than crash reduction. Some of the available studies will be reviewed here, to further point out the problems of using self-reports for driver improvement evaluations.

Conner and Lai (2005) studied the effects of improvement schemes in a sample that was recruited from several counties in England and Wales, comparing them to a control group from areas that did not offer re-education. These were situated mainly in Scotland and Ireland. The method used was self-report, utilizing several scales that are widely used in driver research, amongst them the Driver Behaviour Questionnaire (DBQ; Reason, Manstead, Stradling, Baxter, & Campbell, 1990), the Driver Attitude Questionnaire (DAQ; Parker, Stradling & Manstead, 1996), and a

Sensation Seeking Scale (SSS; Zuckerman, 1994). Also, respondents reported upon their personality, collisions, near misses, and offenses. The last was also acquired from the DVLA for drivers in the education group who supplied their license number. Positive effects were reported for the DBQ scales, but not the attitude scale. The other variables were used as co-variables in the analyses.

The Conner and Lai evaluation is interesting because a lie scale was included, something that is very uncommon in driver research (af Wählberg, 2009). The general finding was that social desirability did not have an effect on the results. However, details were not reported. That an effect of social desirability was indeed present can be suspected, because there were significant decreases in reported errors and lapses of the DBQ, something that would be rather peculiar if the change was in actual behavior, because these 'aberrant driving behaviors' are not considered to be under volitional control (Parker, Lajunen, & Stradling, 1998). The authors of this study did not comment upon this. Indeed, they did not give any reason for why these scales were included in the study at all, so it is not possible to know what they expected in terms of a relation between the educational content and the behaviors these scales are designed to measure. It can also be noted that attitudes and attitude change of traffic violators seem to be unrelated to their crash involvement (Gebers, 1995), which makes the inclusion of the attitude scale in the Conner and Lai study difficult to understand too.

In Meadows (unpublished), the Speed Awareness course given by Lancashire County Council was evaluated using the Driver Behaviour Questionnaire violation scale, and the Driver Attitude Questionnaire. Both these scales showed improvement three months after the course.

Burgess and Webley (1999) also used the DBQ and DAQ scales for evaluation purposes, with similar results and a similar lack of discussion of why they were used, apart from the assumption that a change in attitudes would cause a change in behavior. The same can be said for McKenna (2007), with the difference that the attitude scale was totally unvalidated. The British preference for proxy safety variables is also highly visible in reports that are only peripherally about evaluating safety gains of driver improvement courses (e.g., Davies, Broughton, Clayton, & Tunbridge, 1999; Fylan, Hempel, Grunfeld, Conner, & Lawton, 2006) and for other types of training (e.g., Stanton et al., 2007).

1.4. A new type of driver education

For many years, Thames Valley Police (TVP), as other police forces in England, have run driver re-education courses for drivers who have violated the Highway Code in some way. The education is part of a scheme where the drivers can choose to take a safety course (and pay for it), but avoid a fine and/or some points being added to their records.

In 2008, a new initiative aimed at reducing the collision involvement of young drivers was started by TVP, the Young Driver Scheme (YDS). Under this alternative, drivers below the age of 25 years who are caught for a traffic offense in Thames Valley, by police or camera, are offered to take a safety education that is specifically tailored to this age group; Highway. The content of Highway addresses the specific problems and risks that young drivers meet, like recreational driving, as shown by crash statistics (Clarke, Ward, Bartle, & Truman, 2006).

The course features a workshop, attended by about 20 drivers, under the supervision of a driver trainer, where issues of safety are discussed. Thereafter, each participant does five online modules of education (so-called e-learning, see Appendix), including an assessment of knowledge of the content presented. The course must be concluded within 28 days, with at least four days between each online module. The evaluation of the effects of

the YDS was planned to be an integral part of the scheme, and it was therefore started before the first drivers attended the course.

2. Method

2.1. General

The effects of the YDS course were evaluated by several means, of which the questionnaire part is reported upon in the present paper. The overall goal of using a questionnaire for evaluation purposes was to try to measure behavior more directly than by the use of the sources of collisions and penalty points. These variables may yield indications of the behavior change that safety education is supposed to cause, but as described in the introduction, such variables are also very unreliable. It was therefore an explicit part of the evaluation plan to supplement the self-reports by recorded data on offenses and collisions.

2.2. Questionnaire

The questionnaire was constructed mainly from well known driver inventories; the violations scale from the Manchester Driver Behaviour Questionnaire (DBQ-V; Reason et al., 1990), the (Brief) Driving Anger Scale (DAS; Deffenbacher, Oetting, & Lynch, 1994), the Aggression scale of the Driver Behaviour Inventory (DBI-A; Gulian, Glendon, Matthews, Davies, & Debney, 1988), the (Short) Sensation Seeking Scale (SSS; Slater, 2003), and a three-item drunk driving scale. Included were also single items on the use of cell phones and being fatigued while driving.

The rationale for choosing these inventories included possibilities to compare to previous research on driver improvement (e.g., Conner & Lai, 2005; Edwards, 2005), and some sort of claimed association with crashes. Included in the questionnaire were also items about age, sex, mileage, years of licensed driving, number of collisions since full license, and current number of penalty points.

Finally, the Driver Impression Management (DIM) scale of the Driver Social Desirability Scale (Lajunen, Corry, Summala, & Hartley, 1997) was utilized as a control for socially desirable responding. This scale correlates negatively with self-reported collisions, and very weakly positively with recorded crashes (af Wählberg et al., 2010), as could be expected if it was valid (and self-reports of collisions were influenced by social desirability). It has also been shown for the present data that all driver inventories included, especially the DBQ-V, are strongly correlated with this lie scale, and that this also influences their associations with self-reported crashes (af Wählberg, 2010). For between individuals effects, it has thus been shown that these driver inventories are highly unreliable and prone to common method variance effects. For within individuals biases, no research is known.

2.3. Groups and distribution of questionnaire

The YDS group consisted of drivers below the age of 25 years who had been caught committing a traffic offense (mainly speeding) by a police officer or camera in the Thames Valley area. All such drivers were offered the YDS education instead of a fine and penalty points. About 90% accepted this offer.

To test for changes in behavior of the YDS drivers after the course, a questionnaire was utilized, and delivered online three times: before the start of the course, at the end, and six months after start of the course. The first two waves were integrated into the course, and the first needed to be finished before drivers could enter the modules. The second wave did not carry any such requirement, but a direct link from the last module to the questionnaire probably made the drivers see it as mandatory, although this was not the case. Six months after

the start of the course, each driver was sent an e-mail, asking them to respond to the questionnaire online. This method was expected to result in a very small attrition of respondents for wave two, but a fairly large one for wave three.

To test for various effects not due to the YDS, a control group was constructed for the questionnaire, by the use of an e-marketing scheme. E-mails were sent to possible respondents, asking them to respond to the questionnaire online. All respondents of the control group first wave took part in a sweepstake where five gps systems were given away. After six months, the control group respondents were sent an e-mail asking them to respond to the questionnaire again, indicating another sweepstake.

All start pages for the questionnaire carried information about the respondents being anonymous, that the data would be used for research only, and how to contact the researcher responsible for the evaluation.

2.4. Analyses

The items of each scale within each wave were summed. These scales were tested for changes in means of the scales between waves and between groups, and associations between scales. For all differences, t-tests and effect sizes (Cohen's *d*) were calculated, while associations between items and scales were computed using Pearson correlations.

2.5. Methodological issues

First, the data sets were cleaned up from some obviously false values. For example, one case was deleted because the driver had entered 123 as both number of collisions and penalty points. This case was also suspicious concerning the other values, as it tended to have no variance at all. Also, some values for age and experience were not complete, probably due to errors in the coding process when the data were entered.

Similarly, although the questions about age and licensing used a format of year and month, various analyses showed that some information was not reliable. Thus, some YDS drivers claimed an age above 25 (and thus not eligible for YDS), and having acquired a full drivers license at ages below 17, and even 16 (when disabled persons might have a driving license in the UK). Thirty-seven values were therefore deleted from the age variable, as they were incomplete, or exceeded the allowed value for participation on the YDS course, or indicated ages when the drivers could not have had drivers licenses. Eighty-five values were deleted from the experience variable because they were higher than should be possible given the age restriction of the course, but also incompatible with the age given by the respondent, because they were negative (i.e., the respondent had given a year and month of licensing that was later than the course date), or due to input errors (numbers were not a date).

Originally, the questionnaire used an open question format about mileage per month, where drivers could write in any answer. This format led to unwanted replies in the form of text (such as 'no idea'), as well as clearly erroneous figures, where the response was probably about miles per year instead of per month. The format was therefore changed into 200-mile categories, which could be indicated by clicking on one of these. Sixty-eight values on the mileage variable of the YDS group were deleted. Still, even after cleaning, the data retained on this variable must be seen as very approximate, as few drivers can report correctly upon mileage (af Wählberg, 2009). However, it should be noted that the use that mileage is put to in the present paper was not for individual differences, but as group means, which are fairly reliable (af Wählberg, 2009).

Table 1

The demographic properties of all repeated measurement samples of the YDS and control group, as reported in the first wave. Shown are percent males for sex, thereafter mean/std for age and driving experience in years, miles driven per month, number of collisions and current number of penalty points since licensing for wave 1 and for the last six months for YDS wave 3 and control wave 2. The data in YDS wave 2, and sex, age and experience in wave 3, was not reported in that wave, but in wave 1, and are thus not indicative of any differences in reporting, but of changes in the group that responded.

| Variable | YDS wave 1 | YDS wave 2 | YDS wave 3 | Control wave 1 | Control wave 2 |
|----------------|--------------------|--------------------|-------------------|-------------------|-------------------|
| N | 7601 | 5633 | 729 | 844 | 133 |
| Sex | 60% | 59% | 49% | 33% | 44% |
| Age | 21.7/2.3 | 21.9/2.2 | 22.4/2.2 | 22.2/2.1 | 22.4/2.1 |
| Experience | 3.2/2.1 (N = 7522) | 3.4/2.1 (N = 5583) | 4.0/2.2 (N = 723) | 3.4/2.3 (N = 659) | 3.8/2.5 (N = 103) |
| Mileage/month | 656/561 (N = 7534) | 661/585 (N = 5573) | 614/456 | 441/356 | 423/315 |
| Collisions | 0.58/0.85 | 0.58/0.85 | 0.09/0.39 (6 mo) | 0.33/0.97 | 0.03/0.17 (6 mo) |
| Penalty points | 0.72/1.73 | 0.71/1.67 | 0.19/0.80 (6 mo) | 0.23/1.07 | 0.03/0.26 (6 mo) |

Due to these factors, the N of these demographic variables was somewhat lower than the total for the scales, as the online format resulted in almost complete data for the latter variables, due to it not being possible to skip any items.

The control group featured a number of drivers of older age than those on the YDS. A cut-off was set at <26 years, for two reasons. The number of respondents, especially to the second wave, was not very large, and the YDS drivers included some who claimed to be older than 25 years, the demarcation line for participation on the course. Therefore, the control group was slightly older and more experienced than the YDS group.

The mean time between completion of first and second wave of the questionnaire for the YDS group was 62 days, and between first and third waves 180 days. For the controls, there was a mean difference of 215 days between the two waves.

The rate of response for the second YDS wave was about 80%, and for the third wave about 20%, as compared to the first wave. The reason for the percentages being approximate was that the dates for responding were not exactly fixed, so although the percentage of re-responses could be calculated with a fair degree of accuracy, this number was only indicative of the real percentage, because drivers could be late in responding (i.e., the number of drivers who should have responded at a certain date could only be approximately estimated). Also, these rates are not fully reflected by the present numbers of each wave in the analyses, because at each separate time when data were downloaded, there would be a number of drivers who had not yet

finished the course and been invited to do the second and third waves.

For the control group, only 10 e-mails bounced back when the second wave was sent. The response rate for the rest was 16%.

Descriptive data for the samples can be seen in Table 1. It can be noted that the YDS wave 3 group drivers were fairly similar to the total group who took the course, on all available demographic variables (the differences on age and experience variables were due to differences in dates). They were also similar to the controls regarding age and experience, but different on all other variables, notably crashes, points, and mileage. This finding is in accordance with the conceptualization of the YDS drivers as a high-risk group. However, the mileage difference is important because it signifies that the YDS drivers are not necessarily worse drivers than the average. Actually, the difference in mileage explained most of the difference in reported number of crashes. The difference in number of points, on the other hand, was still large if calculated by mile.

3. Results

3.1. Homogeneity and differences

Two measures of internal consistency of the scales were calculated; Cronbach alpha and the lowest correlation between any items in each scale (see Table 2). The alpha values were acceptable, if not always impressive.

Table 2

Means and standard deviations of the mean response per item for each driver, and Cronbach alphas of the scales, and the lowest correlation between any items of each scale.

| Scale | Value | YDS wave 1 (N = 7639) | YDS wave 2 (N = 5662) | YDS wave 3 (N = 729) | Control wave 1 (N = 847) | Control wave 2 (N = 133) |
|---------------------|----------|-----------------------|-----------------------|----------------------|--------------------------|--------------------------|
| DAS (6 items) | Mean/std | 2.03/0.62 | 2.00/0.72 | 2.17/0.73 | 2.95/0.79 | 2.79/0.72 |
| | Cronbach | .79 | .85 | .84 | .79 | .79 |
| | Lowest r | .31 | .40 | .37 | .30 | .29 |
| DBI-A (5 items) | Mean/std | 1.59/0.50 | 1.71/0.61 | 1.76/0.60 | 2.02/0.74 | 2.16/0.68 |
| | Cronbach | .60 | .69 | .71 | .71 | .70 |
| | Lowest r | .14 | .18 | .20 | .14 | .13 |
| SSS (2 items) | Mean/std | 1.36/0.60 | 1.40/0.68 | 1.46/0.70 | 1.71/0.80 | 1.71/0.72 |
| | Cronbach | 0.77 | .84 | .86 | .82 | .76 |
| | Lowest r | .65 | .73 | .77 | .70 | .67 |
| Drugs (3 items) | Mean/std | 1.07/0.24 | 1.15/0.37 | 1.14/0.33 | 1.15/0.44 | 1.16/0.38 |
| | Cronbach | .57 | .70 | .60 | .80 | .63 |
| | Lowest r | .34 | .45 | .34 | .54 | .41 |
| DBQ-V (7 items) | Mean/std | 1.29/0.35 | 1.33/0.45 | 1.40/0.46 | 1.49/0.57 | 1.44/0.47 |
| | Cronbach | 0.72 | 0.82 | .80 | .84 | .79 |
| | Lowest r | .17 | .29 | .32 | .34 | .18 |
| DIM (7 items) | Mean/std | 3.12/0.91 | - | 2.93/0.93 | 3.05/1.03 | 2.98/1.01 |
| | Cronbach | 0.81 | - | .83 | .87 | .87 |
| | Lowest r | .21 | - | .27 | .31 | .36 |
| Fatigue (1 item) | Mean/std | 2.11/0.88 | 2.15/0.95 | 2.43/0.95 | 2.11/0.93 | 2.25/0.87 |
| Cell phone (1 item) | Mean/std | 1.34/0.67 | 1.37/0.72 | 1.49/0.80 | 1.41/0.80 | 1.29/0.61 |

Table 3
The means and standard deviations of scales in YDS waves 1 and 3, for those who responded to all waves, and the control group. Dependent t-tests calculated for differences between the YDS waves, and independent t-tests between YDS 3 and control, and Cohen's d for both (standard deviation of the left group used as denominator).

| Scale | YDS wave 1, N = 729 | t | d | YDS wave 3, N = 729 | t | d | Control wave 1, N = 847 |
|---------------------|---------------------|---------|-------|---------------------|----------|------|-------------------------|
| DAS (6 items) | 2.03/0.56 | -5.5*** | -0.24 | 2.17/0.73 | -20.2*** | -1.1 | 2.95/0.79 |
| DBI-A (5 items) | 1.61/0.48 | -7.9*** | -0.32 | 1.76/0.60 | -13.5*** | -0.8 | 2.22/0.74 |
| SSS (2 items) | 1.37/0.55 | -4.7*** | -0.18 | 1.46/0.70 | -6.7*** | -0.4 | 1.72/0.81 |
| Drugs (3 items) | 1.07/0.17 | -6.7*** | -0.43 | 1.14/0.33 | -0.7 | 0.0 | 1.16/0.45 |
| DBQ-V (7 items) | 1.30/0.31 | -6.5*** | -0.32 | 1.40/0.46 | -3.4*** | -0.2 | 1.49/0.58 |
| DIM (7 items) | 3.12/0.89 | 6.0*** | 0.21 | 2.93/0.93 | -2.3* | -0.1 | 3.05/1.03 |
| Fatigue (1 item) | 2.21/0.84 | -6.7*** | -0.27 | 2.43/0.95 | 2.1* | 0.1 | 2.11/0.93 |
| Cell phone (1 item) | 1.33/0.64 | -5.4*** | -0.25 | 1.49/0.80 | 2.2* | 0.1 | 1.41/0.80 |

* $p < .05$, *** $p < .001$.

Thereafter, the means and standard deviations for the driver inventory scales and items were calculated. It can be seen that the tendency was for values on all scales and items, except the DIM, to increase with each wave for the YDS group, while the control group remained the same or even decreased slightly. However, the YDS sample had much lower values than the controls in all waves. As the groups in different waves had partial overlap of respondents, no statistical tests were calculated for these differences. However, it can be seen in Table 3 that the same trend was present for the subgroup of respondents of the third YDS wave as for the larger groups in Table 2, and that these differences were highly significant between waves 1 and 3. It can also be seen that all driver inventories had lower values in the YDS group as compared to control wave 1, while the single items of cell phone use and fatigue had opposite effects.

3.2. Stability over time of measurements

Few values of stability over time of driver inventories are available, and these were therefore calculated for the scales used. It can be seen in Table 4 that although these correlations were fairly strong in terms of what is common in social science research, none of them explained more than 50% of the variance, meaning that the drivers' responses to these scales were highly unstable over even such short time periods as the presently used ones.

4. Discussion

The results from the current questionnaire study were in some ways unexpected. Given that the YDS course does have a positive effect on safety, and that drivers report their behavior and reactions truthfully, there should have been a decrease in values with each wave. Finding the opposite therefore makes at least one of these assumptions untenable. However, as the values for YDS were lower than for the control group in all waves, any explanation of the increase over time for the YDS group needs to take this difference into account. Four possible explanations can be forwarded:

- (1) the YDS course changed the drivers' behaviors for worse
- (2) the reporting of the YDS drivers grew more honest with each wave, due to less perceived need to respond in a socially acceptable way
- (3) the behaviors and reactions featured in the driver inventories became more salient with the education, and the drivers were therefore more likely to notice and remember them
- (4) the reporting was too unreliable to estimate any real change

The first explanation runs into a logical problem; why were the values of the YDS drivers lower than those of the control group? If they were true representations of actual behaviors, this would mean that the YDS drivers were actually behaving better than the average driver before the course, which would seem to be very improbable, given that they had been caught for various offenses. If

anything, the YDS drivers could be suspected to be worse drivers than the average.

Second, as the third wave values were lower than those of the control group, it could be the case that the YDS had a positive effect, meaning that the real effect of the course on behavior was offset by the initial artefactual effect on responding. This would explain the differences between groups and waves for the outcome scales. It is also in agreement with the values for the DIM scale, when all groups and waves are considered. In the first wave, YDS respondents lied more than the controls, and had vastly lower values on the other scales, while in the third wave, they lied less than the controls, and still had lower values. This would mean that the drivers were intimidated by their contact with the police, which triggered a strong desire to respond in a socially acceptable way to the questionnaire items. When they had passed the course, this intimidation decreased.

The third explanation is in agreement with the driver inventory results. However, it is in opposition to the fact that the sensation seeking scale changed as the others. The significance of this finding is that sensation seeking is usually considered as a personality trait (Zuckerman, 2007), meaning that it should be fairly stable over time, especially in terms of the mean of a group. Also, the items for sensation seeking are not about driving, and there would therefore seem to be no reason to expect them to change as an effect of the course. In conclusion, this change would rather indicate a response effect (i.e., explanation number two).

The fourth explanation not only acknowledges that there could be response artifacts that have not even been considered here, but also that the various other explanations make the interpretation of the results very difficult. Not only does the logic become a bit complicated, with assumed effects, counter-effects, and differences

Table 4
Correlations of driver inventory scales between waves. Mean time periods were 65 days between YDS waves 1 and 2, and 181 days between waves 1 and 3, and 215 days for the control group. All associations significant at $p < .001$.

| Scale | YDS wave 1 versus 2 (N = 5658) | YDS wave 1 versus 3 (N = 734) | Control wave 1 versus 2 (N = 133) |
|---------------------|--------------------------------|-------------------------------|-----------------------------------|
| DAS (6 items) | .48 | .68 | .68 |
| DBI (5 items) | .47 | .66 | .73 |
| SSS (2 items) | .52 | .73 | .56 |
| Drugs (3 items) | .33 | .56 | .36 |
| DBQ-V (7 items) | .47 | .67 | .67 |
| DIM (7 items) | - | - | .69 |
| Fatigue (1 item) | .47 | .60 | .52 |
| Cell phone (1 item) | .41 | .56 | .51 |

between samples, but it can also be questioned whether the DIM scale accurately captures the assumed socially desirable responding that could explain the results.

Of these various explanations, the second one would seem to have the best backing in the present data; drivers feel less need to present themselves in a favorable light when they have finished the course. This interpretation is also supported by the findings of a very similar study for seatbelt wearing education, where the same kind of effect as in the present study was found, but without the problem of an adverse effect in the variables estimating the effect of the course (af Wählberg, submitted for publication). On all scales measuring concepts that were unrelated to the course content, reports were worse after the course, while for the personal seatbelt wearing items, there was an improvement. Similarly, Kroner and Weekes (1996) reported that prison inmates near the end of their term scored lower than those at the beginning on a lie scale. However, the last explanation cannot really be ruled out, a conclusion that underscores the need for objectively recorded data to be used for further evaluation of the YDS.

It can be noted that the present results indicate strong effects that are not due to the intervention (i.e., response artifacts between situations). As noted previously, the reporting of the drivers is unreliable when individual differences measures are calculated (af Wählberg, in press). However, strong effects of socially desirable responding between situations have not been reported previously, probably because this has rarely been studied in real situations (af Wählberg, 2009; see also the intriguing results of Falk, 2010). In essence, it would seem like the presently used, very popular driver inventories are very susceptible to social desirability effects between situations, at least when the situation encountered is socially sensitive, as can be expected from an encounter with the law that results in a punishment.

The main conclusions that could be drawn from the present results were therefore that questionnaires are difficult to use as outcome measures in socially sensitive situations, and that for the present evaluation, any interpretation of the self-reported data need corroboration from objective sources, which for drivers would be recorded offenses and collisions. Such data are being gathered for the YDS course and will be reported upon in future papers.

The present results are in strong disagreement with those of Burgess and Webley (1999), Conner and Lai (2005), and Meadows (unpublished), who reported improvement on the DBQ and the DAQ. No explanation of this difference has been found. It could be suspected that the amount of intimidation experienced would be a function of age, but the associations between age and the scales included were all very weak, so the difference in age groups between studies do not seem to be a possible cause. Another explanation could be differences in the exact circumstances of the delivery of the questionnaire, although this can not be ascertained.

It is interesting to note that in many cases of driver improvement, the reduction of offenses is seen as a positive outcome in itself, and re-offending used as the most important criterion for course effects (e.g., Davies, Harland, & Broughton, 1999). It is thus very apparent that the ultimate goal of reducing collisions is often seen as an automatic effect of an effect on the criterion used, despite evidence to the contrary, as stated in the introduction of this paper. Another facet of this proxy variable thinking is that offenses are thought of as strong indicators of individual differences in risk for crashes, as for example explicitly stated by the UK notion of 'High Risk Offenders' (Davies, Broughton, et al., 1999).

The online format of the questionnaire used in the present study is still not common for traffic safety research, despite the popularity of driver inventories and the strong increase of all web activities in the last few years. It can therefore be of some interest to note a few of the advantages and drawbacks of this methodology. First, the automatic conversion of the answers into

downloadable data makes the format extremely simple and cheap when large samples are wanted. When the web pages containing the items and the database have been constructed, data are gathered without any additional work at all, regardless of the number of respondents.

However, it must also be acknowledged that the online format may have drawbacks that have not yet been recognized. For example, when the only way to move forward to the next item is to give an answer to the preceding one, it is possible that the replies will be forced from subjects who do not have a ready answer, and who, in another format, would have chosen to not give an answer at all. It would therefore seem to be important to always add a 'don't know' box.

The use of questionnaires in driver safety research has for decades been seen as unproblematic, and hundreds of studies have been undertaken using such instruments, without any kind of testing of the validity of the results (af Wählberg, 2009). It would seem that there is a need for methodological development within this area of research, whether it be individual or within-subjects differences. First, a lie scale needs to always be included, preferably one that has been validated for driver research. Second, other common method variance effects in questionnaires need to be studied. Today, there is virtually no knowledge about how strongly results within traffic safety research are influenced by mechanisms such as self-generated validity (Feldman & Lynch, 1988).

Acknowledgements

This study was part of the evaluation of the Young Driver Scheme, instigated by Thames Valley Police, DriveTech and a2om. Several people within these organizations helped out with data, services, information and a nice work climate. For the present study, the most important contributors were Chris Johnson (a2om), who set up the online questionnaire, and Dan Campsall (TVP), who organized the e-marketing scheme. Malcolm Collis (TVP) ably supervised the project until his retirement. Credit is also due to Mick Doyle (TVP), who made it all happen. An anonymous reviewer provided useful feedback.

Appendix A. Highway e-learning

This on-line educational package consists of five modules, which are described below. The material is largely visual and inter-active, with the general set-up being an animated scenario where you are driving a car and end up in some sort of incident. This is re-played several times from different angles, with risk factors pointed out. The student can also go back and study any part of the module at will.

In most of the modules, questions are asked about what the student thinks about the risks involved when the scenario has been presented. Thereafter, the risks are pointed out. Afterwards, the answers given by the student are compared to the correct ones.

Finally, an assessment is carried out, where the student has to reply correctly to twenty out of twenty-five questions about the content of the module. If this limit is not achieved, the student has to re-do the module.

Anatomy of a crash: You are driving along a road with your friends when the phone rings; you pick it up and crash into a tree. Thereafter, three outcomes of crashes is described; what happen to the car, what happen to the humans inside, and what happen to the family and friends of the victim(s). This module has a spoken track, while the rest only have text, and no assessment at the end.

Attitude and Alertness: You are following a friend's car to a party, and your mates want you to keep up with it. An incident occurs at a roundabout.

Safety Margins: The concept of the safety bubble is discussed and examples of this provided, including stopping distances and skidding.

Overtaking: Uses the scenario of overtaking a bus, discusses the hazards involved, and asks whether it is necessary, legal and safe in this situation. Other situations are also discussed, including overtaking on a dual carriageway.

Anticipation and Hazard Perception: Discusses the hazards in a scenario, and scanning techniques for anticipating, spotting and reacting to them, taking into account your reaction time and blind spots.

References

- Burgess, C., & Webley, P. (1999). Evaluating the effectiveness of the United Kingdom's National Driver Improvement Scheme. In G. Grayson (Ed.), *Behavioural Research in Road Safety IX* (pp. 39–54). Crowthorne: Transport Research Laboratory.
- Carcallion, L. I., Rachid Salmi, L., & Atout-Route Evaluation Group. (2005). Evaluation of a program to reduce motor-vehicle collisions among young adults in the county of Landes, France. *Accident Analysis and Prevention*, 37, 1049–1055.
- Christie, R. (2001). The Effectiveness of Driver Training as a Road Safety Measure: A Review of the Literature. *Report 01/03*. Melbourne: RACV.
- Clarke, D. D., Ward, P. J., Bartle, C., & Truman, W. (2006). Young driver accidents in the UK: The influence of age, experience, and time of day. *Accident Analysis and Prevention*, 38, 871–878.
- Conner, M., & Lai, F. (2005). Evaluation of the Effectiveness of the National Driver Improvement Scheme. *Road Safety Report No. 64*. London: Department for Transport.
- Davies, G. P., Broughton, J., Clayton, A., & Tunbridge, R. (1999). The High Risk Offender Scheme for Drink-drivers. *TRL Report 394*. Crowthorne: Transport Research Laboratory.
- Davies, G. P., Harland, G., & Broughton, J. (1999). Drink/driver Rehabilitation Courses in England and Wales. *TRL Report 426*. Crowthorne: Transport Research Laboratory.
- Deffenbacher, J. L., Oetting, E. R., & Lynch, R. S. (1994). Development of a driving anger scale. *Psychological Reports*, 74, 83–91.
- Edwards, I. (2005). An analysis of the National Driver Improvement Scheme by referral type. In L. Dorn (Ed.), *Driver Behaviour and Training, Volume II, Second International Conference on Driver Behaviour and Training, Edinburgh 15-17 November, 2005* (pp. 19–32). Aldershot: Ashgate.
- Falk, B. (2010). Do drivers become less risk-prone after answering a questionnaire on risky driving behaviour? *Accident Analysis and Prevention*, 42, 235–244.
- Feldman, J. M., & Lynch, J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology*, 73, 421–435.
- Ferdun, G. S., Peck, R. C., & Coppin, R. S. (1967). The teen-aged driver: An evaluation of age, experience, driving exposure and driver training as they relate to driving record. *Highway Research Record*, 163, 31–53.
- Fylan, F., Hempel, S., Grunfeld, B., Conner, M., & Lawton, R. (2006). *Effective Interventions for Speeding Motorists*. London: Department for Transport.
- Gebers, M. A. (1995). Knowledge and Attitude Change and the Relationship to Driving Performance among Drivers Attending California Traffic Violator School. *RSS-95-147*. Sacramento: California Department of Motor Vehicles.
- Gulian, E., Glendon, A. I., Matthews, G., Davies, D. R., & Debney, L. M. (1988). Exploration of driver stress using self-reported data. In J. A. Rothengatter, & R. A. de Bruin (Eds.), *Road User Behaviour: Theory and Research* (pp. 342–347). Maastricht: van Gorcum.
- Hakkert, A. S., Gitelman, V., Cohen, A., Doveh, E., & Umansky, T. (2001). The evaluation of effects on driver behavior and accidents of concentrated general enforcement on interurban roads in Israel. *Accident Analysis and Prevention*, 33, 43–63.
- Harano, R. M., & Peck, R. C. (1972). The effectiveness of a uniform traffic school curriculum for negligent drivers. *Accident Analysis and Prevention*, 4, 13–45.
- Janke, M. K. (1994). Mature driver improvement program in California. *Transportation Research Record*, 1438, 77–83.
- Johnson, H. M. (1939). Evidence for educational value in drivers' "clinics." *Psychological Bulletin*, 36, 674–675.
- Johnson, H. M., & Cobb, P. W. (1938). The educational value of "drivers' clinics." *Psychological Bulletin*, 35, 758–766.
- Kaestner, N. F. (1968). Research in driver improvement: The state of the art. *Traffic Quarterly*, 23, 497–520.
- Ker, K., Roberts, I., Collier, T., Beyer, F., Bunn, F., & Frost, C. (2005). Post-license driver education for the prevention of road traffic crashes: A systematic review of randomized controlled trials. *Accident Analysis and Prevention*, 37, 305–313.
- Klein, D. (1966). A reappraisal of the violation and accident data on teen-aged drivers. *Traffic Quarterly*, 20, 502–510.
- Kroner, D. G., & Weekes, J. R. (1996). Balanced inventory of desirable responding: Factor structure, reliability, and validity with an offender sample. *Personality and Individual Differences*, 21, 323–333.
- 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.
- Lonero, L. P. (2008). Trends in driver education and training. *American Journal of Preventive Medicine*, 35, S316–S323.
- Lund, A. K., & Williams, A. F. (1985). A review of the literature evaluating the defensive driving course. *Accident Analysis and Prevention*, 17, 449–460.
- Masten, S. V., & Peck, R. C. (2004). Problem driver remediation: A meta-analysis of the driver improvement literature. *Journal of Safety Research*, 35, 403–425.
- Mayhew, D. R., Simpson, H. M., Williams, A. F., & Ferguson, S. A. (1998). Effectiveness and role of driver education and training in a graduated licensing system. *Journal of Public Health Policy*, 19, 51–67.
- McGuire, F. L., & Kersch, R. C. (1969). A study of history, philosophy, research methodology, and effectiveness in the field of driver education. *University of California Publications in Education*, 19, Berkeley: University of California Press.
- McKenna, F. (2007). Do attitudes and intentions change across a speed awareness workshop? *Behavioural Research in Road Safety XVII* (pp. 265–276).
- Meadows, M.L. (unpublished). *Evaluation of Lancashire County Council's Speed Awareness Course*.
- 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.
- Nasvadi, G. E. (2007). Changes in self-reported driving behaviour following attendance at a mature driver education program. *Transportation Research Part F*, 10, 358–369.
- Nasvadi, G. E., & Vavrik, J. (2007). Crash risk of older drivers after attending a mature driver education program. *Accident Analysis and Prevention*, 39, 1073–1079.
- Owsley, C., McGwin, G., Jr., Phillips, J. M., McNeal, S. F., & Stalvey, B. T. (2004). Impact of an educational program on the safety of high-risk, visually impaired, older drivers. *American Journal of Preventive Medicine*, 26, 222–229.
- Parker, D., Lajunen, T., & Stradling, S. G. (1998). Attitudinal predictors of interpersonally aggressive violations on the road. *Transportation Research Part F*, 1, 11–24.
- Parker, D., Stradling, S. G., & Manstead, A. S. (1996). Modifying beliefs and attitudes to exceeding the speed limit: An intervention study based on the Theory of Planned Behavior. *Journal of Applied Psychology*, 26, 1–19.
- Planek, T. W., Schupack, S. A., & Fowler, R. C. (1974). An evaluation of the national safety council's defensive driving course in various states. *Accident Analysis and Prevention*, 6, 271–297.
- Podsakoff, P. M., Mackenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903.
- Reason, J., Manstead, A., Stradling, S., Baxter, J., & Campbell, K. (1990). Errors and violations on the roads: A real distinction? *Ergonomics*, 33, 1315–1332.
- Roberts, I., Kwan, I., & Cochrane Injuries Group Driver Education Reviewers. (2008). School-based driver education for the prevention of traffic crashes. (Review). *The Cochrane Library*, 2008(1).
- Schmitt, N. (1994). Method bias: The importance of theory and measurement. *Journal of Organizational Behavior*, 15, 393–398.
- Slater, M. D. (2003). Alienation, aggression, and sensation-seeking as predictors of adolescent use of violent film, computer and website content. *Journal of Community*, 53, 105–121.
- Stanton, N. A., Walker, G. H., Young, M. S., Kazi, T., & Salmon, P. M. (2007). Changing drivers' minds: The evaluation of an advanced driver coaching system. *Ergonomics*, 50, 1209–1234.
- Struckman-Johnson, D. L., Lund, A. K., Williams, A. F., & Osborne, D. W. (1989). Comparative effects of driver improvement programs on crashes and violations. *Accident Analysis and Prevention*, 21, 203–215.
- af Wahlberg, A. E. (2009). *Driver Behaviour and Accident Research Methodology; Unresolved Problems*. Aldershot: Ashgate.
- af Wahlberg, A. E. (2010). Social desirability effects in driver behaviour inventories. *Journal of Safety Research*, 41, 99–106.
- af Wahlberg, A.E. (submitted for publication). Evaluation of an e-learning seatbelt wearing intervention.
- af Wahlberg, A. E., Dorn, L., & Kline, T. (2010). The effect of social desirability on self reported and recorded road traffic accidents. *Transportation Research Part F*, 13, 106–114.
- af Wahlberg, A.E., Dorn, L., & Kline, T. (in press). The Manchester Driver Behaviour Questionnaire as predictor of road traffic accidents. *Theoretical Issues in Ergonomics Science*. doi:10.1080/14639220903203376.
- Zuckerman, M. (1994). *Behavioural Expressions and Biosocial bases of Sensation Seeking*. Cambridge: Cambridge University Press.
- Zuckerman, M. (2007). The sensation seeking scale V (SSS-V): Still reliable and valid. *Personality and Individual Differences*, 43, 1303–1305.

Anders af Wahlberg graduated from Uppsala University (Sweden) with a B. A. in history (1994) and an M. A. in psychology (1996). He thereafter became a research assistant at the Center for Risk Research, Stockholm School of Economics (1996–1998), but turned to driver research and earned his PhD in psychology (2006), at the Department of Psychology, Uppsala University. He currently holds a researcher position at the same department. Main research interests: Driver behavior, traffic accidents, methodology, most often together.