Evaluation of an e-learning seatbelt wearing intervention

A. E. af Wåhlberg
Department of Psychology
Uppsala University
P. O. Box 1225
751 42 Uppsala
Sweden
e-mail: anders.af_wahlberg@psyk.uu.se
Homepage: http://www.psyk.uu.se/hemsidor/busdriver
Tel: +46-18-471 25 90, +46-18-33 90 95
Fax: +46-18-471 21 23

Abstract

Problem: Potential traffic offenders do not always seem to be deterred by fines. Therefore, education has often been tested as an alternative to punishment, but without proven effects. Advances in information technology and education have, however, made it possible to deliver education on line, with a type of content that could have more effect than previous curricula.
Method: An e-learning course for drivers caught not wearing their seatbelts was evaluated, using an on-line questionnaire, which was distributed before and three months after the course. Several controls for changes in socially desirable responding between situations (before/after) were used.
Results: The seatbelt items indicated that seatbelt wearing had changed in a positive direction after the course, while scales measuring other behaviors changed for the worse. Also, less faking was indicated by a lie scale, explaining why most reports had a negative change.
Discussion: These results show that there was an effect of social desirability between situations (there was less of it after the course, as compared with before), but that despite this counter-effect, a positive change was noted for the items where this benefit was expected as a result of the education. This indicated that the course had had the desired effect. However, this interpretation of the results is limited by the self-report method as well as strong self-selection of the second wave respondents.
Impact on industry: The use of driver education delivered on line would seem to be a preferable alternative to fines and other punishments for driving offences, although more research is needed on the effects of this mode of delivery. The cost-effectiveness is probably high, but requires a separate investigation. Evaluation studies using questionnaires cannot rely solely on poorly validated scales, as is currently common practice, but need to apply controls for social desirability. Preferably, other sources of data should be used instead.

Key words: e-learning, seatbelt, driving offence, traffic safety, driver education

1. Introduction

Within traffic safety, seatbelts are probably one of the most important inventions ever implemented (e.g. Hakkert, 1969; Evans, 1986). Simple, cheap and reliable, there is really only one drawback; they rely heavily upon drivers and passengers of vehicles to actually use them. It has also been claimed that the rate of pedestrian casualties increases with their use (Conybeare, 1980), but this question does not seem to have been resolved. Although wearing rates are rather high in industrialized countries (e.g. Streff & Wagenaar, 1989; Parada, Cohn, Gonzalez, Byrd & Cortes, 2001), there are still a few percent of people who do not use them, and these individuals may to some degree also be the people who would especially need them, due to their driving style (Evans, 1987; Hunter, Stewart, Stutts & Rodgman, 1993; Stewart, 1993; Cooper, 1994; Dee, 1998; Vassallo et al., 2007), although not all studies have concluded this (Wasielewski, 1984).

Therefore, the road users who are non-users of seat belts are an important group to reach, to change their behavior. This has traditionally been done by police surveillance and fines. However, although there seem to be a considerable effect of making seatbelt-wearing mandatory (Stewart, 1993), as well as of enforcement (Williams, Wells, & Lund, 1987; Mortimer, Goldsteen, Armstrong & Macrina, 1990; Salzberg & Moffat, 2004; Shults, Elder, Sleet, Thompson, & Nichols, 2004; Houston & Richardson, 2005), there are still, as noted, people who do not use them. These road users are apparently not deterred by the risk of social disapproval and/or a fine and complimentary strategies are therefore needed.

Education and driver training have been the preferred alternatives to punishment, with training involving actual driving, while education refers to more theoretically oriented approaches, where the aim is to change behavior by influencing how the driver thinks and feels. When evaluating training and education (most often called 'driver improvement'), it has been concluded that only very small, if any, positive effects have been found, and often the opposite, on the target behavior (see reviews by Kaestner, 1968; Lund & Williams, 1985; Struckman-Johnson, Lund, Williams & Osborne, 1989; Janke, 1994; Masten & Peck, 2004; Ker, Roberts, Collier, Beyer, Bunn & Frost, 2005). Although these reviews have included in their material several methods that can be seen as rather different from each other, for example warning letters and group educational sessions, what can be seen as more pure educational efforts (informing about risks) have so far not been shown to have any beneficial effects (see the review in af Wåhlberg, 2010a).

In the United Kingdom, driver improvement schemes are common. Instead of paying a fine, drivers can elect to take a course instead (and pay for it). These are usually classroom based, and aimed at minor offences. Several evaluations have been undertaken of these schemes through the years. For example, Broughton, Buckle and Pearce (2005) investigated re-offending rates at a national level, and in general found no difference between scheme attendees and first-time offenders that had not taken a course. It was not clearly stated how the control group had been recruited. A self-selection bias can therefore not be excluded.

On the other hand, Davies and Smith (2003) found a fair effect on re-offending rates over several years for drink driving courses, while several studies using self-reported data also reported improvement (e.g. Conner & Lai, 2005; Meadows, unpublished; Burgess & Webley, 1999). The latter studies and others were discussed in detail in af Wåhlberg (2010a) and found not to be reliable, due to methodological shortcomings and uncertain reporting.

The conclusions from the UK experience of driver improvement schemes would therefore seem to be similar to the international (mainly American) one; different reports yield different results, with the main possible explanations for these discrepancies being differences in evaluation methodology and course contents. The possibility of some courses being effective therefore remains, although the probability would seem to be slight.

However, the strong development within information technology during the last decades has made it possible to present educational material in a different way from previous methods, both regarding sound, movement and interactivity, but also for the setting, where students can now take a course in their own home, something that might increase their compliance. Thus, on-line education for drivers has recently been introduced in the UK, with one of the target populations being traffic offenders. Within this group, several different sub-types of offenders have been identified and educational material developed. One of these, the Young Driver Scheme (YDS), has been evaluated and would seem to carry some promise of positive effects for re-offending (af Wåhlberg, submitted), while the questionnaire results were difficult to interpret (af Wåhlberg, 2010a). The present study investigated the effects of an e-learning scheme on seat-belt wearing behavior of drivers who had committed such an offence in the Thames Valley Police region of the UK.

Unfortunately, objective data on drivers' behavior and accident records is difficult to acquire, for different reasons. Although many different facets of driver behavior can be measured today (speed, acceleration, fuel consumption etc), a system where this kind of information can be extracted for driving offenders does not exist. Also, even when recorded crash and offence data (which are commonly used as dependent variables in driver improvement evaluations) can be used, these have the disadvantages of low variance and a fair degree of randomness over the short time periods for which educational efforts can be expected to have an impact. Whether a driver crashes or is caught for an offence is dependent upon many factors beside his/her own behavior, and accident record data is therefore a crude estimate of any changes in what they do. For these reasons, many researchers today turn to the self-report method as a way to measure behavior change. For the present type of behavior change, crashes would also not be an appropriate outcome variable, and self-reports would seem to be the only possible alternative.

However, the use of self-reports in driver research is always under the threat of various types of response bias, which might lead to artefactual effects between or within individuals (af Wåhlberg, 2009; 2010b; af Wåhlberg, Dorn & Kline, 2011). In particular, socially desirable responding seems to be a strong biasing factor in studies on differences in driver behavior and safety (af Wåhlberg, 2010a; 2010b). Any study on a socially sensitive topic that uses self-report needs therefore take very good care to counter such bias.

That seat belt wearing is indeed a socially sensitive subject in some countries probably explains why people consistently over-report its use (Streff & Wagenaar, 1989; see also the review in af Wåhlberg, 2009). However, it should also be noted that an evaluation is different from an individual differences study, because each individual can be compared to him/herself over time, with the same method. The question that arises then is whether there might be situational factors that can bias the evaluation, creating effects that are not due to the intervention under investigation.

In a previous evaluation of an e-learning scheme, it was found that there was a massive effect of social desirability in the responses between questionnaire waves (af Wåhlberg, 2010a). The values for various driver behavior inventories were in fact

higher (worse behavior) after the course than before, when honest reporting would have yielded at least no difference, and a negative difference if the course had had the desired effect. However, the increase in self-reported violations, aggressive acts etc was interpreted as an effect of social desirability in the pre-course measurement, with respondents becoming more honest after they had taken the course, and thus had atoned for their offence.

This interpretation of the effects was possible because the values of the course group were consistently lower than those of a control group (which should not have been possible, as the offender group was a sample selected for its bad behavior), and because the values on a lie scale changed in this direction. However, most informative was the change on a scale that had little to do with the course content; sensation seeking. This characteristic is viewed as a personality trait (Roberts, Walton & Viechtbauer, 2006; Zuckerman, 2007), meaning that it should only change over rather long time periods.

It can therefore be suspected that the perceived reporting situation for offenders is very much dependent upon whether the authorities still have some sort of legal claim on the participant. Such effects have been found in other socially sensitive situations, such as personnel selection (Barrick & Mount, 1996). This type of effect is therefore a very significant problem for evaluations using questionnaire methods linked to socially sensitive situations, as the socially desirable responding effect between waves tends to swamp any real effect, or increase it. However, this interpretation of the phenomenon would seem to be novel, as is the report of the phenomenon itself, and it would therefore need to be repeated with different questionnaire scales.

It can be noted that this effect (the increase of self-reported dangerous driving behavior after a driver improvement course) could not have been due to self-selection among the YDS participants, as they were under the impression that the questionnaire was a mandatory part of the course, yielding almost a hundred percent reply rate for the first wave, and for the second wave eighty percent.

In the present study, the apparently new method of controlling for situational effects of social desirability by inclusion of personality scales was used again, but with some further developments. Basically, scales were included that could be considered driving and/or safety-related, but where the content was not covered by the e-learning course and the construct as such could be considered not to be subject to quick changes, while the self-reports of these constructs could be expected to be influenced by social desirability.

To sum up, the present study set out to evaluate the effects on seat belt wearing behavior due to an on line driver improvement course for drivers caught not wearing their seatbelt in Thames Valley Police district, UK. The method used was an on line questionnaire with several different controls for situationally induced socially desirable responding.

2. Method

2.1 General

Road users caught without their seatbelt in Thames Valley Police region (UK) were offered the opportunity of taking an on-line improvement course (fee £25) instead of paying a fine (£60). The aim of the study was to evaluate whether the course had any effect on seatbelt-wearing behavior.

The evaluation was carried out using a two-wave questionnaire method, distributed three months apart. The first wave questionnaire was delivered as part of the on-line

module, meaning that when drivers received their login details, they were first directed to the questionnaire, and were requested to respond to this before starting the online module. The second wave was distributed by sending an e-mail to each participant three months after the course, asking them to complete the questionnaire online. The questionnaire was displayed on a single page, and the respondents could therefore go back and change their answers.
No control group was utilized in the present study. This was due to two factors (apart from the practical problems involved). In the YDS evaluation, a fairly large, random control sample had been gathered, using an e-marketing scheme. It showed no effect over the same time period (six months) as between the second and third waves of the YDS questionnaire (which yielded strong effects), for various behavior scales.
For this reason, and because some of the scales (conscientiousness and sensation seeking) in the present study measured constructs that are considered to be stable over time (i.e. personality; Roberts, Walton & Viechtbauer, 2006; Zuckerman, 2007), any difference found could be expected to be due to the situation and/or the education. As the course content was not related to the control scales, this should not influence them, and if so, in the opposite direction of the expected effect of situationally induced social desirability.
For the seat belt items, a similar argument could be made. If a change in general seatbelt-wearing rate co-occurred with the course, this could create an artefactual result in the present study. However, as seatbelt wearing was already high[1] in this area (Walter, Narine & Buttress, 2009), and no other initiatives for increasing it apart from the course under evaluation were known, this would seem improbable.

2.2 Samples and questionnaire waves
The offenders were divided into active drivers of vehicles (at least once every month) and non-drivers, according to their response on a question put by the company distributing the login details. The latter group (about seven percent of the total) responded to the same questionnaire as the drivers, but with the driving-related items removed. This data is not treated in the present analysis, as the available second wave sample was too small.
The first wave of the questionnaire was distributed before the e-learning course. Three months after the respondents had finished the course they were sent an e-mail and asked to respond again. The same questionnaire was used in both waves.

2.3 Scales and items
The questionnaire consisted of demographic items (see Table 1), questions on sensation seeking, conscientiousness, social desirability, seatbelt wearing (own and other's), and beliefs about seatbelts. Age, years of licensing, penalty points and mileage was reported in categories (of five, five, three and 200, respectively), and these measures were therefore rather approximate. However, as drivers cannot report upon their driving with any high degree of accuracy (af Wåhlberg, 2009), it was considered that no information was lost in this way. The reason for using categories was the on-line format, where it had previously been found that the use of free text windows resulted in non-numerical responses (e.g. dunno!), which were difficult and time-consuming to convert to numbers. Responses to all other items were reported on

[1] More than ninety percent of drivers and front seat car occupants used seatbelts, while values were lower for taxi drivers and passengers.

five-step scales, to the questions ‘How much do you agree?’ ('Not true' to 'Very true') or 'How often do you?’ ('Never' to 'All the time').
A two-item version of the Sensation Seeking Scale (SSS; Slater, 2003) was included, as it had been used in the YDS evaluation. There, it had been found to have very good homogeneity (af Wåhlberg, 2010a), and a test-rest correlation of about .50.
The Conscientiousness scale was chosen as the main control for situational social desirability, as it is considered as the most easily faked dimension of the Big Five (Furnham, 1997; McFarland & Ryan, 2000). To some degree, this has been corroborated by research (Mersman & Shultz, 1998). Also, it has been found to change very little over time, in terms of mean values (Roberts, Walton & Viechtbauer, 2006). Any change in mean value over three months time would therefore indicate a bias in responses. The two conscientiousness items were taken from Gosling, Rentfrow and Swann (2003), and slightly adapted to fit into the general format of the present questionnaire. Short forms of the Big Five have been shown to have acceptable reliability and predictive validity (Langford, 2003).
The Driver Impression Management (DIM, seven items) scale of the Driver Social Desirability Scale (Lajunen, Corry, Summala & Hartley, 1997) was utilized as a further control for situational differences in socially desirable responding. This scale has been validated in terms of being negatively correlated with self-reported collisions, and very weakly positively associated with recorded crashes (af Wåhlberg, Dorn & Kline, 2010), which means that it at least captures one type of driving-related faking, the biased reporting of crashes.
Four items on driver errors from the Manchester Driver Behaviour Questionnaire (DBQ) were taken from Kontogiannis (2006). The violation scale of the DBQ has previously been shown to be very susceptible to socially desirable responding (af Wåhlberg, 2010a; 2010b).
Five seatbelt items were included from Gras, Cunill, Sullman, Planes and Font-Mayolas (2007), two of which had been found to discriminate between drivers who had been observed with and without seatbelt, thus validating these items with objective data. To these were added one item about the likelihood of a crash in relation to seatbelt wearing. These items were used to form two scales; seatbelt beliefs and other’s belt wearing behavior.
Finally, one question was asked about how often the respondents actually wore a seat belt, and one item about forgetting to put the seatbelt on. These items were analyzed separately.

2.4 Analysis
The respondents who had completed the questionnaire twice were compared between waves on the two seat belt items (own wearing and forgetting to put it on) and beliefs about seatbelts, which were the only measures that were supposed to show any effect of the improvement course. All other scales were expected to function as situational bias controls, indicating whether the respondents became more honest in their responses after the course.
The scales for conscientiousness, errors and sensation seeking were therefore not seen as valid indicators of actual behavior, but as measurements of faking, whether intentional or not, and thus similar to the lie scale included. This is a very different way of using these scales, which have ordinarily been accepted as valid, without much evidence for this assumption (af Wåhlberg, 2009). However, assuming that most of the differences found on these scales are actually due to reporting biases mean it is possible to explain results that would otherwise be very intriguing.

It was expected that respondents would report less conscientiousness and more sensation seeking and errors after the course, while faking less. For the seatbelt wearing items, it was expected that there would be a positive effect of the course, and a negative effect of the situation, but which effect would be stronger was not possible to predict. For the scale 'Other's seatbelt wearing', it could not be directly influenced by the course, and the content was not considered socially sensitive, and no change would therefore seem logical. However, educated drivers might be more inclined to try to make their family and friends to use seatbelts, or simply influence them by example, and a small positive effect could therefore result.
All differences were tested with dependent t-tests, and Cohen's d was calculated.

2.5 Educational material
The on-line educational material consisted of some twenty web pages, where the main parts were based upon a scenario where the driver is in the car with friends, running late and colliding with another vehicle. The impact of a crash on the human body at the stated speed was described, with and without seatbelt. There were also questions about how often and why the offender wears a seat belt, and statistics about the safety effect of seatbelts.
The course participants could move through the pages at their leisure, and return to previous ones. At the end of the module, there were fifteen assessment questions, ten of which the respondent needed to get right to pass the course.

3. Results

About seventy percent of the offenders accepted the offer of the improvement course. All of these responded to the first questionnaire wave. The re-response rate was 6.3 percent in the second wave.
Descriptive data for the drivers who responded to both waves can be seen in Table 1. For comparisons between those who only responded to the first wave and the others, see af Wåhlberg and Poom (in preparation).

Tables 1-2 about here

In Table 2, the responders to wave 2 were compared between waves. Here, it was evident that the drivers tended to report worse values after the course, on most variables that were not included in the course curriculum. The drivers' own seatbelt wearing and forgetting rates were changed in the positive direction, as were their perceptions of other's seatbelt wearing. No change in their beliefs about seatbelt wearing took place.
Given the assumption that drivers were affected by the situation and therefore responded in a less socially desirable manner after the course (as indicated by the DIM lie scale), it could be expected that the intercorrelations between variables would decrease (the common method variance effect within each wave was decreased). Therefore, the means of all associations between the control scales were calculated for both waves (see Table 3). The outcome measures were not included, as they had been manipulated by the course. In ten out of fifteen comparisons, the correlation was weaker after the course, with the mean correlations being .266 and .241, respectively. This a difference of eighteen percent in amount of variance explained.

Table 3 about here

4. Discussion

The evaluation results would seem to indicate an effect of the course on the target behavior; safety belt wearing, over a period of three months. That this effect comes through despite the counter-effect of social desirability between situations, as indicated by the control scales, is further evidence that there was a strong shift in behavior in this sample. On the other hand, the driver's beliefs about seatbelt wearing were not influenced, something which is at odds with the course content.

The use of self-reported seatbelt data is a limitation of the present analysis. As has been shown in several studies, drivers' reports of their seatbelt wearing are not very reliable, and systematic biases (Parada, Cohn, Gonzales, Byrd & Cortez, 2001) and environmental effects exist (Fhanér & Hane, 1973). However, given the application of several controls for socially desirable responding, the results can probably still be trusted, for this sample.

The sample self-selection is another, and stronger, limitation of the present results, which is made worse by the low response rate. In principle, it could be argued that the second wave responders were different (more easily influenced, or worse to begin with) from the other offenders on the course. However, as all course takers responded to the first wave of the questionnaire, comparisons can be made between these two groups. The same situation applies to the previous YDS evaluation, and the ongoing red light scheme project. These data sets have been pooled and analyzed in a separate paper (af Wåhlberg & Poom, in preparation). In general, it was found that differences between later wave respondents and others were rather small.

For the seatbelt data, the strongest effects found were for age and experience (t-values around 5), with those who responded to both waves of the questionnaire having higher values. The general impression would seem to be that the second wave responders were (or wanted to appear to be) more cautious drivers than the non-responders. They might therefore have tended to be drivers who were somewhat more easily influenced by the course content, as compared to the others. However, as they started out as slightly better cases, they would not have as much room for improvement as the others.

Another kind of limitation of the present study was that the population under study was of a special kind (traffic offenders), and it is therefore uncertain whether the results can be applied to other road user groups (i.e. it is not certain whether e-learning courses can successfully applied to basic education of road users).

It should also be noted that no control group drawn from the same population as the intervention group was utilized in the present study. Unfortunately, it is difficult to randomly assign offenders to intervention and control conditions in the present type of study. This is due to legal reasons, but also because drivers who were left without any sort of legal response would in many ways constitute a poor comparison. The proper control group would therefore be drivers who received a fine for their offence, without the opportunity to participate in a course instead. Such a group could only be found in another police district, or recruited from before the implementation of the driver improvement scheme. None of these alternatives could be created in the present study. However, if being caught for a seatbelt offence had the effect of increased seatbelt wearing, this effect should have started before the course. Similarly, the offenders paid their fee before responding to the questionnaire, so this should not have had any effect on the present results either.

The finding of lower correlations between scales after the course was a novel result. Although it does not tell us how much of the variance is due to reporting bias, at least it tells us the lowest possible figure (about a fifth of the variance explained in individual differences), and is a further methodological tool for those who are interested in testing their data for such biases, in this case between differently socially sensitive situations.

The inclusion of various behavior scales as controls for situation-induced socially desirable responding would seem to be a new method, at least within traffic safety research. The basic logic behind their use is that there is no reason to expect any change to occur during the measurement period, apart from that caused by the situation, because there is nothing in the course curriculum that touches upon these aspects. Furthermore, Conscientiousness and Sensation Seeking are considered personality traits, and those are, by definition, stable over time. The present effect on those two variables can be compared to those found in a meta-analysis on mean changes in personality traits over time by Roberts, Walton and Viechtbauer (2006). The largest d was 0.26, for a ten-year period, in a positive direction. In the present data, the conscientiousness scale changed in the negative direction, although not significantly so, while sensation seeking did increase significantly.

The DBQ error scale was similarly affected by the situation as the personality scales. This is noteworthy not only because such a result has never been reported before, but also because it indicates that driving error reporting is susceptible to reporting bias. That behavior of this type is socially sensitive (as indicated by the correlation with the DIM lie scale) is interesting. Apparently, behavior that is not under volitional control can also be seen as undesirable to report, possibly due to it being socially embarrassing.

It is very noteworthy that not only were there sizeable effects of the situation on most variables, but apparently, self-reports can be affected, even by fairly mild 'threats'. In the present study, drivers were not in a particularly threatening situation, but in their own homes or other private contexts, atoning for a not very serious traffic offence, which many of their peers would probably see as quite acceptable.

The present results, where the results on some self-report scales that are often used in driver research were found to be strongly influenced by the situation, are complimentary to previous findings concerning individual differences in responding, where biased reporting and common method variance effects have been detected (af Wåhlberg, 2010a; 2010b; af Wåhlberg, Dorn & Kline, 2010; 2011). Given these strong effects, many studies can be questioned as to the validity of their findings. This goes for all studies using the DBQ that have not employed common method variance controls, and self-reported seatbelt wearing (e.g. Chliaoutakis, Gnardellis, Drakou, Darviri & Sboukis, 2000). Similarly, conclusions regarding the efficacy of various methods for behavior change based upon self-reported data without controls are dubious as to their validity (e.g. the review Fylan, Hempel, Grunfeld, Conner & Lawton, 2006).

The change in the scale used for reporting seatbelt wearing was 0.35 on a five-step scale, an increase of eight percent. What could this transfer to in terms of a percent increase in using the safety belt for these individuals? Unfortunately, no technique for transforming scales with qualitative anchors into ratio level scales seems to be known. What we do know is that the end points represent zero and a hundred percent seatbelt wearing, respectively.

Given the limitations of the present study, as detailed, the results would still seem to indicate that at least some drivers' behavior can be changed in the short term,

regarding their seat belt wearing, via an online improvement course. These results are in agreement with the YDS evaluation, but without the ambiguity of worse results on the target items after the course. The present results therefore support the interpretation of the YDS data in af Wåhlberg (2010a).
However, the most important supporting evidence would be had if objective data could yield a positive effect of an e-learning course. Such data has been analyzed for the YDS (offences), but only for rather limited samples (af Wåhlberg, submitted). The results would seem to support the current conclusion of a positive effect of e-learning on driver behavior. Also, an evaluation of an e-learning course for red light offenders is under way, utilizing a questionnaire of the same type as the present one, as well as offence data. Given that all these studies take place as planned, and yield similar results, a case could be presented for the positive effects of the kind of online educational material that was used in the presently studied course.

Acknowledgement
The online questionnaires and the e-mail scheme were set up by Chris Johnson (a2om), which is gratefully acknowledged. Also, Claire Benson, Bryan Pritchard and Malcolm Collis from Thames Valley police provided information.

5. References

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81,* 261-272.

Broughton, J., Buckle, G. S., & Pearce, L. M. (2005). *The Effects of the National Drivers Improvement Schemes.* TRL Report 649. Crowthorne: Transport Research Laboratory.

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.

Chliaoutakis, J. E., Gnardellis, C., Drakou, I., Darviri, C., & Sboukis, V. (2000). Modelling factors related to the seatbelt use by the young drivers of Athens. *Accident Analysis and Prevention, 32,* 815-825.

Conner, M., & Lai, F. (2005). *Evaluation of the Effectiveness of the National Driver Improvement Scheme.* Road Safety Report No. 64. London: Department for Transport.

Conybeare, J. A. (1980). Evaluation of automobile safety regulations: The case of compulsory seatbelt legislation. *Policy Sciences, 12,* 27-39.

Cooper, P. J. (1994). Estimating the over involvement of seat belt non-wearers in crashes and the effect of lap/shoulder restraint use on different crash severity consequences. *Accident Analysis and Prevention, 26,* 263-275.

Davies, G. P., & Smith, L. R. (2003). *Re-convictions of Drink/Drive Course Attenders: A Six Year Follow-Up.* TRL Report TRL 574. Crowthorne: Transport Research Laboratory.

Dee, T. S. (1998). Reconsidering the effects of seat belt laws and their enforcement status. *Accident Analysis and Prevention, 30,* 1-10.

Evans, L. (1986) The effectiveness of safety belts in preventing fatalities. *Accident Analysis and Prevention, 18,* 229-241.

Evans, L. (1987). Belted and unbelted driver accident involvement rates compared. *Journal of Safety Research, 18,* 57-64.

Fhanér, G., & Hane, M. (1973). Seat belts: the importance of situational factors. *Accident Analysis and Prevention, 5,* 267-285.

Furnham, A. (1997). Knowing and faking ones five-factor personality score. *Journal of Personality Assessment, 69,* 229–243.

Fylan, F., Hempel, S., Grunfeld, B., Conner, M., & Lawton, R. (2006). *Effective Interventions for Speeding Motorists.* London: Department for Transport.

Gras, M. E., Cunill, M., Sullman, M. J., Planes, M., & Font-Mayolas, S. (2007). Predictors of seat belt use amongst Spanish drivers. *Transportation Research Part F, 10,* 263-269.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big Five personality domains. *Journal of Research in Personality, 37,* 504-528.

Hakkert, A. S. (1969). The economic savings from wearing safety belts. *Accident Analysis and Prevention, 1,* 153-158.

Houston, D. J., & Richardson, L. Jr. (2005). Getting Americans to buckle up: The efficacy of state seat belt laws. *Accident Analysis and Prevention, 37,* 1114-1120.

Hunter, W. W., Stewart, J. R., Stutts, J. C., & Rodgman, E. A. (1993). Observed and self-reported seat belt wearing as related to prior traffic accidents and convictions. *Accident Analysis and Prevention, 25,* 545-554.

Janke, M. K. (1994). Mature driver improvement program in California. *Transportation Research Record, 1438,* 77-83.

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.

Kontogiannis, T. (2006). Patterns of driver stress and coping strategies in a Greek sample and their relationship to aberrant behaviors and traffic accidents. *Accident Analysis and Prevention, 38,* 913-924.

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.

Langford, P. H. (2003). A one-minute measure of the Big Five? Evaluating and abridging Shafer's (1999a) Big Five markers.

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.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85,* 812–821.

Meadows, M. L. (unpublished). *Evaluation of Lancashire County Council's Speed Awareness Course.*

Mersman, J. L., & Shultz, K. S. (1998). Individual differences in the ability to fake on personality measures. *Personality and Individual Differences, 24,* 217-227.

Mortimer, R. G., Goldsteen, K., Armstrong, R. W., & Macrina, D. (1990). Effects of incentives and enforcement on the use of seat belts by drivers. *Journal of Safety Research, 21,* 25-37.

Parada, M. A., Cohn, L. D., Gonzalez, E., Byrd, T., & Cortes, M. (2001). The validity of self-reported seatbelt use: Hispanic and non-Hispanic drivers in El Paso. *Accident Analysis and Prevention, 33,* 139-143.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132,* 1-25.

Salzberg, P. M., & Moffat, J. M. (2004). Ninety five percent: An evaluation of law, policy, and programs to promote seat belt use in Washington state. *Journal of Safety Research, 35,* 215-222.

Shults, R. A., Elder, R. W., Sleet, D. A., Thompson, R. S., & Nichols, J. L. (2004). Primary enforcement seat belt laws are effective even in the face of rising belt use rates. *Accident Analysis and Prevention, 36*, 491-493.

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.

Stewart, J. R. (1993). Seat-belt use and accident involvement - a comparison of driving behavior before and after a seat-belt law. *Accident Analysis and Prevention, 25,* 757-763.

Streff, F. M., & Wagenaar, A. C. (1989). Are there really shortcuts? Estimating seat belt use with self-report measures. *Accident Analysis and Prevention, 21,* 509-516.

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.

Vassallo, S., Smart, D., Sanson, A., Harrison, W., Harris, A., Cockfield, S., & McIntyre, A. (2007). Risky driving among young Australian drivers: Trends, precursors and correlates. *Accident Analysis and Prevention, 39,* 444-458.

Walter, L. K., Narine, S., & Buttress, S. (2009). *Thames Valley Seat Belt Survey 2009.* Published Project Report PPR 416. Crowthorne: Transport Research Laboratory.

Wasielewski, P. (1984). Speed as a measure of driver risk: Observed speed versus driver and vehicle characteristics. *Accident Analysis and Prevention, 16,* 89-102.

Williams, A. F., Wells, J. K., & Lund, A. K. (1987). Shoulder belt use in four states with belt use laws. *Accident Analysis and Prevention, 19,* 251-260.

af Wåhlberg, A. E. (2009). *Driver Behaviour and Accident Research Methodology; Unresolved Problems.* Aldershot: Ashgate.

af Wåhlberg, A. E. (2010a). Re-education of young driving offenders; effects on self-reports of driver behavior. *Journal of Safety Research, 41,* 331-338.

af Wåhlberg, A. E. (2010b). Social desirability effects in driver behavior inventories. *Journal of Safety Research, 41,* 99-106.

af Wåhlberg, A. E. (submitted). Re-education of young driving offenders; effects on recorded offences and self-reported collisions.

af Wåhlberg, 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 Wåhlberg, A. E., Dorn, L., & Kline, T. (2011). The Manchester Driver Behaviour Questionnaire as predictor of road traffic accidents. *Theoretical Issues in Ergonomics Science, 12,* 66-86.

af Wåhlberg, A. E., & Poom, L. (in preparation). An empirical test of non-response bias in surveys.

Zuckerman, M. (2007). The sensation seeking scale V (SSS-V): Still reliable and valid. *Personality and Individual Differences, 43,* 1303-1305.

Table 1: Descriptive results for the drivers who responded to both questionnaire waves. Shown are the percentage of males and the mean and standard deviation of age, experience (years of license), current number of penalty points on record, mileage per month, and the number of culpable and non-culpable crashes for the last three years. All data self-reported. Age, experience, points and mileage were reported in categories of five years, and the numbers presented were calculated by taking the mid-point of each category.

Variable	Sex	Age	Experience	Points	Mileage	Culpable crashes	Non-culpable crashes
Value	83.8%	41.2/14.5	21.0/14.0	0.87/1.66	817/597	0.138/0.398	0.184/0.504

Table 2: Descriptive data and difference tests for the group that responded to both waves of the questionnaire. Dependent t-tests and Cohen's d calculated for the differences between the waves. N=505. Number of items for each scale within parentheses.

Variable	Conscientiousness (2)	Sensation seeking (2)	DBQ errors (4)	DIM (7)
Wave 1	4.56/0.58	1.49/0.75	1.37/0.43	3.66/0.89
Wave 2	4.51/0.62	1.64/0.83	1.45/0.50	3.55/0.94
t	1.6	-5.1***	-3.7***	3.1**
d	0.09	0.20	-0.19	0.12
Variable	Seatbelt beliefs (5)	Seatbelt use (1)	Other's seatbelt use (2)	Forgetting seatbelt (1)
Wave 1	1.51/0.61	4.40/0.67	1.29/0.57	1.95/1.16
Wave 2	1.52/0.61	4.75/0.56	1.35/0.63	1.53/0.89
t	-0.1	-11.7***	-2.3*	7.4***
d	0.02	-0.52	-0.11	0.36

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

Table 3: The correlations between waves for each scale (first column, N=505) and the intercorrelations between scales in each wave. In the upper, right-hand corner, wave 1 (N=8013), in the lower left hand corner wave 2 (N=505).

Scale	Between waves	Conscientiousness	Sensation seeking	DBQ errors	DIM	Seatbelt beliefs	Other's seatbelt use
Conscientiousness	.383***	-	-.204***	-.234***	.257***	-.244***	-.220***
Sensation seeking	.604***	-.231***	-	.224***	-.338***	.293***	.266***
DBQ errors	.503***	-.263***	.102*	-	-.265***	.208***	.236***
DIM	.602***	.246***	-.383***	-.184***	-	-.276***	-.222***
Seatbelt attitude	.549***	-.144***	.274***	.181***	-.236***	-	.421***
Other's seatbelt use	.384***	-.155***	.267***	.257***	-.191***	.335***	-

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