Department for Transport

TRL THE FUTURE OF TRANSPORT

PUBLISHED PROJECT REPORT PPR2010

Driver2020 - an evaluation of interventions designed to improve safety in the first year of driving

Report D2: Effectiveness of interventions delivered to learner drivers

J Weekley, S Helman, H Makosa, N Harpham, J Hutton

DRIVER 2020

Report details

Report prepared for:	Department for Transport
Project/customer reference:	1-170
Copyright:	© TRL Limited
Report date:	April 2024
Report status/version:	6
Quality approval:	Lucy Fuller **(Project Manager)** Graham Grayson **(Technical Reviewer)**

Disclaimer

This report has been produced by TRL Limited (TRL) under a contract with the Department for Transport. Any views expressed in this report are not necessarily those of the Department for Transport.

The information contained herein is the property of the Department for Transport. Whilst every effort has been made to ensure that the matter presented in this report is relevant, accurate and up-to-date, TRL Limited cannot accept any liability for any error or omission, or reliance on part or all of the content in another context.

Acknowledgements

Firstly, TRL would like to thank the Department for Transport for funding and commissioning this research. Secondly, we would like to thank the delivery partners who assisted in the design of the interventions, and who delivered them during the Driver2020 trial; these were The Royal Society for the Prevention of Accidents, The Driving Instructors Association, Trak Global, DriverMetrics, and Agilysis. Further thanks are due to the following individuals who assisted in the design of the trial: Dr Graham Grayson, Professor Frank McKenna, Professor Steven Stradling, Dr Lisa Dorn, Professor Nicola Christie, Professor Jane Hutton, Professor Johnathon Ehsani, Professor Alan Tapp, Professor Mark Horswill, and Dr Willem Vlakveld. Finally, the following organisations served on the stakeholder panel that contributed insights around how the trial might best support policy: The Welsh Government, Transport Scotland, BRAKE the road safety charity, National Highways, The Parliamentary Advisory Council for Transport Safety, The National Fire Chiefs Council, National Police Chiefs Council, RAC Foundation, Road Safety GB, The Driver and Vehicle Standards Agency, The Driver and Vehicle Licensing Agency, Transport for London, and The Association of British Insurers.

Table of Contents

Executive summary

Background

It is a known challenge in road safety that young and novice drivers are at a greater risk of involvement in collisions than more experienced drivers, for reasons associated with their younger age, and their relative lack of on-road experience. A large evidence base confirms this (see Appendices A.1 and A.2).

Historically there have been two broad approaches to reducing collision risk for young and novice drivers. The first is to use the licensing system to directly target age and experience factors, using legislation to set a minimum age for licensure or minimum periods of time for people to spend in the supervised learning phase, and to set restrictions on higher risk driving situations after solo driving begins. There is a strong evidence base showing it can be effective at reducing risk (see Appendix A.3) in other jurisdictions. One post-test measure in the Great Britain system is the two year probationary period during which new drivers on their first full licence who accumulate six or more penalty points have their licence revoked and are required to go through the learning process again. The second broad approach is to use non-legislative measures such as driver training and education to try to equip drivers with better skills and knowledge to keep safe. There is less evidence that this approach is effective (see Appendix A.4).

The Department for Transport (DfT) commissioned TRL to conduct the Driver2020 project, which sought to build evidence for educational, training and technology-based non-legislative approaches by evaluating the most promising of these in a randomised controlled trial reflecting a potential real-world voluntary rollout of such measures.

The Driver2020 project

In a previous DfT-funded review of evidence, including a consultation with stakeholders, Pressley et al. (2016) identified five interventions that showed promise, and were the most suitable for trialling in a research programme committed to in the 2015 Road Safety Statement. Three of these were designed to be delivered during the learning period, before people had passed their practical driving test. These interventions were:

- **Logbook:** an app designed to prompt learners to obtain more on-road practice with instructors or with other supervising drivers such as parents, and to have this practice cover as wide a range of situations as possible
- **Hazard perception training:** an online set of three hazard perception training modules designed to improve hazard perception skill
- **Education intervention:** a whole-day classroom-based intervention (and later due to the impact of COVID-19, an eLearning intervention) in which attendees take part in several activities designed to equip them with knowledge and skills, and ongoing self-monitoring strategies, to make them safer as drivers.

This report presents findings from the learner driver arm of the DfT-funded Driver2020 study, in which the effectiveness of these three interventions was evaluated. Three other reports present findings on two additional interventions delivered to novice drivers, findings from qualitative research with users and providers to understand engagement with the interventions, and a summary report for the whole project (Weekley et al.2024a; Hitchings et al., 2024; Helman & Weekley, 2024).

The two research questions for this element of the evaluation were (for each learner intervention):

- How effective is the intervention at reducing collisions in the first 12 months of post-test driving?
- Does engaging with the intervention lead to other changes in relevant surrogate measures? (Surrogate measures are outcomes used as substitutes for collisions either because they are known to be related to collision risk – for example speeding – or they reflect things that are targeted by an intervention).

Method

Both arms of the study used a randomised encouragement design. In the learner arm, this meant that participants who registered their interest were randomly assigned to one of four groups, with three providing the opportunity to engage with an intervention (logbook, hazard perception training, or education), and one being a control group that was offered no intervention. One specific requirement for the project was that participants should not be provided with overly generous incentives to engage with the intervention to which they were assigned; incentives needed to be seen as sustainable in any wider roll-out, meaning they were very modest; in the learner arm only the logbook intervention had any specific incentive, with participants in this group having an opportunity to receive learner driver insurance at a discounted rate.

The learner arm of the study was advertised to learner drivers using two main methods. The first was a leaflet provided in the pack accompanying the provisional driving licence. The second was through Approved Driving Instructors (through the Driving Instructors Association and through other routes) telling their clients about the study. Most (94%) of the 16,214 learner participants who registered their interest in the trial were recruited via the former method, with the remainder finding out through the latter, or through other means.

In order to be eligible for the study, participants needed to be aged between 17 and 24 years when they began learning to drive, and either have spent no more than ten hours driving so far or intend to begin learning to drive within the next two months. They were also required to be at least intending to have access to a vehicle for driving when they passed their test. If assigned to an intervention group, participants were put in touch with their intervention provider following registration. All participants were also asked to complete a survey when they passed their practical driving test which asked them about their experience learning to drive, and further surveys 3, 6 and 12 months after their test pass date. These later surveys asked about their experience driving post-test, about collisions they had been involved in, and about several other surrogate measures of safety, some of which were related specifically to outcomes targeted by the interventions. Participants were given a £5 retail voucher (or equivalent charity donation if they preferred) for each survey they completed and were entered into a prize draw if they wished, to win a year's car insurance and other prizes such as retail vouchers and tablet computers.

Recruitment of learner participants began in January 2019 and continued until early March 2020. Therefore, there were a number of ways in which learning to drive, and post-test driving experience were impacted by the COVID-19 pandemic and associated lockdowns and restrictions. These factors were considered and controlled where possible in the analysis.

Analysis

When all participants had completed their 12 months post-test period, two analyses were run to understand the effectiveness of the interventions.

First, the study investigated how effective each intervention was at reducing collisions in the 12-month post-test period; this was assessed by comparing the average number of self-reported collisions in each intervention group with that of the control group. This analysis is sometimes referred to as an 'intention to treat' analysis; it included all participants (for whom there were survey data) in each group, regardless of whether they engaged with the intervention. Such an approach aims to understand what would happen at the population level in real-world conditions, rather than looking specifically at people who follow the treatment protocol as intended.

The second analysis is known as a 'per-protocol analysis' – so-called as it includes only those people who followed the treatment protocol as intended – so looking only at the impact of the intervention on those people who were known to engage to some degree. This analysis used the surrogate measures as outcomes, since the sample sizes were much too small to permit any robust analysis of collision numbers. Because the numbers of people who engaged were low, the study compared relevant surrogate measures – which matched the intended outcomes of the interventions – between the intervention and control groups.

For the logbook intervention, engagement was defined as using the app and recording at least one session. For the hazard perception training intervention engagement was defined as completing at least one of the three modules; for this group, which had a higher engagement rate than other interventions, it was also possible to define a 'high engager' group (completed two or more modules) and still have a reasonable sample size at each time point in the study. Finally for the education intervention, engagement was defined as attending an in-person course or completing at least one module online.

Limitations

Several limitations to the study design should be considered when reading the findings and conclusions.

- The data were gathered during the COVID-19 pandemic, which would have affected learning and post-test driving. Generalisability to future circumstances in which there is no global pandemic may be affected.
- The sample is biased towards females and is an opportunity sample, which limits generalisability.

- The self-selecting nature of the sample (both in terms of registering for the trial and in terms of providing survey data) means that the findings cannot necessarily be generalised to all learner drivers in Great Britain. However, the design (randomised allocation to groups) does mean that self-selection bias should not play any role in group differences.
- Most data are self-reported meaning some social-desirability effects may be present and data reflect the interpretation of participants to survey questions.
- The engagement with interventions was at a very low level, meaning the findings might not reflect what would happen with greater engagement.
- Some groups had smaller-than-planned samples, meaning that some very small effects may have been missed due to lower statistical power.
- The findings related to surrogate measures provide us with evidence of the behaviour change that results from the interventions as these analyses only include those who engaged; however, the very low numbers engaging with some interventions mean that some very small effects may have been missed.
- The findings only cover the first year of driving and cannot be extended beyond this period (for example we cannot say whether behaviour would be impacted beyond this period, or whether any changes seen would persist).

Findings

The findings from the study are reported below, first for the effect of the interventions on collisions, and then on surrogate measures. It should be noted that the COVID-19 pandemic affected both learning to drive, and post-test driving during the study for the majority of participants. Thus, some caution should be exercised in how generalisable the findings are to future circumstances.

Collisions

- Greater mileage was associated with more collisions.
- As drivers accumulated on-road experience the increased risk associated with increased mileage reduced; in other words, although mileage overall increased risk, the increase was not linear.
- The COVID-19 pandemic also led to changes in collision risk (after controlling for mileage). Those who passed their test before the pandemic were more likely to be involved in collisions than those who passed after.

The findings from the intention to treat analysis were:

- There was no statistically significant difference between the logbook (N=916), hazard perception training (N=769), education (N=682), and control (N=925) groups on the main outcome measure of self-reported collisions in the first 12 months of post-test driving.

- This was true after controlling for mileage driven and for the impacts of the COVID-19 pandemic, age, and gender.
- Although the numbers of participants were much smaller than had been originally planned for the learner arm (less than half of the target sample) there was almost no difference in the collision rate of the groups and none of the statistical tests were close to reaching statistical significance.

The lack of an effect on collisions is not surprising given the low engagement rates, since even if the interventions were extremely effective for those who engaged, they would only be having an effect on a small proportion of the intervention group; an analogy is that if a medicine exists that cures a disease with 100% efficacy, and only 1% of people take the medicine, then the real world effectiveness of the medicine will be 1%. Unfortunately, it is not possible to examine differences in collisions solely in the people who engaged with the interventions, as the samples sizes were insufficient to support such an analysis; this is why surrogate measures were used for the per-protocol analysis (see below).

Surrogate measures

The lack of differences in collisions between groups lends greater importance to the per-protocol analyses, which look only at those people known to have engaged to some degree with the interventions, on a range of surrogate measures targeted for change by those interventions. It needs to be noted that the engagement 'dose' or 'amount' of intervention exposure was low even in those people who did engage, and it may be that the effects seen were limited by this.

The focus is on statistically significant findings, however due to the very low sample sizes in the engager groups, consideration was also given to any findings that might potentially have reached statistical significance had the sample sizes been as planned. This analysis showed the following effects for those who engaged with the interventions (all findings relative to the control group):

For the logbook intervention:

- Engagers reported a statistically significantly larger proportion of their learning to drive as being with passengers in the car (20% for engagers, N=84, versus 14% for the control participants, N=410). (Safety impact unclear)
- Engagers reported a non-significant difference in the total number of hours of practice during learning (81.8 for engagers, N=84, versus 76.3 for control participants, N=410). (Safety impact unclear)

For the hazard perception training intervention:

- Engagers reported a statistically significantly lower frequency of driving above the speed limit in their first three months of post-test driving (14.3 for engagers, N=263, versus 11.4 for control participants, N=445, where '0' means 'not at all' and '100' means 'all the time'). (Would likely improve safety)

- Engagers had a statistically significantly lower number of attempts to pass the theory test (1.19 for engagers, N=318, versus 1.37 for control participants, N=696). (Compatible with higher hazard perception scores, which would likely improve safety)
- Engagers (N=300) reported driving styles that were non-significantly less ‘inattentive, careless, irresponsible and risky’ and less ‘irritable, impatient and intolerant’ than control participants (N=549). (Would likely improve safety)

For the education intervention:

- Engagers reported a statistically significantly lower proportion of learning to drive as being in the dark (16% for engagers, N=79, versus 21% for control participants, N=416). (Safety impact unclear.)
- Engagers reported a statistically significantly higher proportion of learning to drive as being on dual carriageways (17% for engagers, N=79, versus 15% for control participants, N=416). (Safety impact unclear.)
- Engagers reported a non-significant lower proportion of learning to drive as being with passengers in the car (13% for engagers, N=79, versus 14% for control participants, N=416). (Safety impact unclear.)
- Engagers reported a non-significant higher frequency of driving post-test as involving speeding (17.8 for engagers, N=82, 15.6 for control participants, N=440, where 0 is never exceed the speed limit, and 100 is exceed the speed limit all the time) . (Safety impact unclear.)

Conclusions

The learner driver arm of the Driver2020 project aimed to test the real-world effectiveness of three interventions on both collision involvement, and a number of surrogate safety measures, in the first 12 months of post-test driving.

The sample size was much lower than had been planned for (due to the impact of COVID-19 on the number of people who were able to pass their test in time to contribute data to the study). This meant that the statistical power in the study (the study’s chances of detecting a defined reduction in collisions) was lower than anticipated. Nonetheless there was almost no difference in the observed rate of collisions in the intervention and control groups, and none of the statistical tests were near reaching statistical significance. Therefore, the findings from this sample suggest that if any of these interventions were offered on a similar voluntary basis as in this study to learner drivers aged 17-24 in Great Britain, they would be unlikely to lead to any measurable reduction in collision risk.

An important finding that helps in the interpretation of this main conclusion is that when such interventions were offered under conditions of voluntary engagement, levels of engagement were extremely low.

When only participants who had engaged with the interventions were considered, all three interventions were associated with statistically significant changes in relevant surrogate measures, relative to the control group.

Further work with the Driver2020 dataset could help further elucidate the effects of the COVID-19 pandemic on learning to drive and on early post-test driving in young and novice drivers. It also provides a recent dataset that can be used to examine this group, and the long-understood road safety challenge they present.

1 Introduction

1.1 The challenge of young and novice drivers

This section summarises the known road safety challenge presented by young and novice drivers, and a short description of research with this group including the programme of applied work delivered in Great Britain throughout the 1990s and 2000s, that laid the foundations for the approach taken in the Driver2020 project. It is a summary of the more detailed background provided (with references) in Appendix A.

The conclusion that emerges from this evidence is that young and novice drivers have an elevated collision risk for reasons associated with their youthfulness and their inexperience. In short, younger drivers and less-experienced drivers are at greater risk. Successfully reducing risk requires interventions that act on one or both areas.

Approaches that bring about an increase in the age at which someone becomes licensed have reduced collisions. The same is true of approaches that provide drivers with greatly increased levels of on-road experience before licensure. Finally, approaches that limit exposure to the riskiest situations in early driving after licensure – allowing experience to build up in lower risk situations before access to higher risk situations is granted – are effective in reducing collisions. In Great Britain, the Road Traffic (New Drivers) Act 1995 mandates a two year probationary period during which new drivers on their first full licence who accumulate six or more penalty points have their licence revoked and are required to go through the learning process again. Two years after passing their first driving test, or on passing again after having their licence revoked under the Road Traffic (New Drivers) Act 1995, drivers enter the full licence stage. Note that driver licensing is devolved in Northern Ireland and so this project covers Great Britain only.

Many interventions based on education, technology and training have not been developed and delivered in an evidence-led way. Currently, there are no standards or guidance for road safety education, so interventions can vary hugely in quality, delivery and content. They have delivered poor results because they either target things that may not be relevant for safety, or they target things that are relevant but do so inadequately. An example of the former is the traditional 'skills' approach to driver training. In the past this has failed to bring about safety improvements due to a focus on specific vehicle control skills such as 'recovering from skids' on the assumption that improving such skills will reduce collision risk. An example of potentially inadequate targeting can be found in attempts to provide drivers with knowledge about risk on the assumption that this knowledge will lower drivers' collision risk by bringing about changes to their driving behaviour. The assumptions underlying these approaches have turned out to be incomplete. Vehicle control skills do not appear to be adequate to ensure safety, and changing behaviour requires more than just a provision of information about risk. Approaches that have fared better include hazard perception training, which focuses on the skill of anticipating potential hazards on the road ahead, although the literature on such approaches is still relatively immature in so far as being able to demonstrate a direct link from the training to collision outcomes.

Some training and education approaches have even been associated with increased crash risk. There are plausible mechanisms that can explain how this can happen. For example, it is believed that some skid training courses have led to some drivers taking additional risks, as they assume that their new training will keep them safe; in reality the skills fade quickly without practice, and therefore this confidence is unfounded. Some education approaches have resulted in people gaining access to driving earlier (and therefore at a younger age) than would otherwise have been the case, putting them at more risk as a result.

In light of the relative lack of success of many existing interventions based on education, technology and training, even in graduated licensing systems research attention is now revisiting such approaches such that they can support other policy approaches. The hope is that if interventions can be based on sound behavioural science, and can be focused on the right mechanisms, they can add value.

1.2 Origins of the project

In the 2015 Road Safety Statement (DfT, 2015) the Government committed to "Undertaking a £2 million research programme to identify the best possible interventions for learner and novice drivers" (p8). The Statement also set the context for this research programme by stating the following in paragraphs 1.11 and 1.12:

> "1.11: Ten years ago, there were fewer options for reducing the elevated collision risk within the young driver population. Many foreign governments placed legislative 'graduated driver licencing' restrictions on their young people. These options include restricting driving to the hours of daylight or not allowing the carriage of passengers, for months or even years after passing tests."
>
> "1.12: Technology is one of the ways that we can help young drivers be safer. Technology is now emerging that can manage novice driver risk in a more bespoke way without restricting the freedoms of all of our young people. In short, there are modern and sophisticated non-legislative alternatives that treat each young driver as an individual with their own distinct risk profile."

The work that underpinned the commitment to the research programme that became the Driver2020 project was therefore focused on finding non-legislative interventions that were best suited for trialling.

1.3 New approaches to reducing collision risk

The year after the publication of the Road Safety Statement, the Department for Transport (DfT) commissioned a review of interventions for young and novice drivers, focusing specifically on identifying interventions based on education, training and technology that showed promise either theoretically (for example being based on accepted models of learning or behaviour change) or based on evaluation data on some relevant surrogate measure of risk (for example driving behaviour) (Pressley et al., 2016). The intention was that these interventions could then be evaluated in a controlled trial in Great Britain, looking specifically at their effectiveness in reducing collisions.

The brief for the Pressley et al. project specifically excluded approaches that are legislative in nature and focused on non-legislative options.

In this way the Pressley et al. review was the work that laid the ground for the Driver2020 project; the intention was that the candidate interventions identified would subsequently be tested in a large-scale controlled trial, using collisions as the main outcome measure.

Pressley et al. concluded that there were seven approaches that showed promise. Three of these involved parental engagement in the learning and early driving of young and novice drivers, three involved the use of telematics or app-based approaches to monitor and manage driving risk, and one was hazard perception training. All were found to show promise either based on existing evidence of changes in relevant surrogate measures or based on linking to a known risk factor for young and novice drivers in a theoretically coherent and plausible way.

1.4 The Driver2020 project

Following a stakeholder workshop on the feasibility of the interventions identified in Pressley et al. (2016), four were recommended for trialling.

Two of these were designed to be applied in early post-test driving ("novice driver interventions"):

- **Mentoring agreements:** a set of materials for use by novice drivers and mentors (for example parents) in voluntarily setting restrictions on early post-test driving (for example driving with peer-age passengers, or driving at night)
- **Telematics:** an app-based intervention designed to provide feedback to novice drivers on their driving style, with various incentives provided for safer driving.

Two were designed to be applied during the learning stage ("learner driver interventions"):

- **Logbook:** an app designed to encourage more on-road practice, covering a broader range of driving conditions and road types, during the learning period
- **Hazard perception training:** a set of three e-learning modules designed to improve hazard perception skill.

Following the recommendations from Pressley et al., a decision was also taken by the Department for Transport to include an education-based intervention for trialling (to be applied during the learning period). The reasoning for this was that, despite there being limited evidence uncovered in the Pressley et al. review, delivery capacity for this type of intervention exists in Great Britain, and there was already widespread delivery. Driver2020 provided an opportunity to develop an education intervention informed by behaviour change techniques and targeted at learner drivers. It was therefore decided that if strong content, informed by behaviour change theory, could be designed, then evaluating this alongside the other approaches would be of value:

- **Classroom-based education:** a whole-day classroom-based intervention in which attendees take part in several activities designed to equip them with knowledge and skills, and ongoing self-monitoring strategies, to make them safer as drivers.

The Driver2020 project was therefore a national trial to evaluate the effectiveness of these five interventions. The project was procured in 2017 and was registered at the ISRCTN registry (https://www.isrctn.com/ISRCTN16646122).

1.5 Research questions

The main research question for each intervention was as follows:

- How effective is the intervention at reducing collisions in the first 12 months of post-test driving?

A secondary research question for each intervention was:

- Does engaging with the intervention lead to other changes in relevant surrogate measures?

For all interventions, due to the fact that participants were free to choose to engage, the following research question was also asked:

- What were the factors that led people to engage with the intervention, or the barriers that stopped them engaging?

1.6 This report

This report presents the findings from the learner driver arm of the Driver2020 project. Four reports cover the whole project:

- D1 – Effectiveness of interventions delivered to novice drivers (Weekley et al., 2024a). Presents analysis and findings from the quantitative evaluation of the novice driver interventions.
- D2 (this report) – Effectiveness of interventions delivered to learner drivers (Weekley et al., 2024b). Presents analysis and findings from the quantitative evaluation of the learner interventions.
- D3 – Delivery of interventions and engagement by novice and learner drivers (Hitchings et al., 2024). Presents analysis and findings from the qualitative evaluation using interviews with participants and delivery partners for both novice and learner interventions.
- D4 – Summary of findings (Helman & Weekley, 2024). Overall project report summarising the key findings from each part of the trial.

Please note that these four references are cross-referenced in the reports (including this one) when it is useful to do so; however, they have been published at the same time as each other as part of the Driver2020 project, rather than being part of the wider existing literature. A supplementary appendix containing the data collection surveys and intervention logic models is also available (Weekley & Helman, 2024).

This report is structured as follows:

Section 1 has described the background to the Driver2020 project – covering its origins within policy, the aim of the project and the research questions. Section 2 describes the overall method for the study in terms of its high-level design and approach. The impact of the COVID-19 pandemic on the trial is also discussed in this section. Note that much of Sections 1 and 2 are included in all project reports, with adjustments relating to the focus of the report (for example in this document – COVID-19 impact is described mainly in terms of its impact on the learner arm).

Sections 3, 4 and 5 then discuss the learner driver interventions only. Section 3 describes the method used in the learner driver arm of the study, including the detailed design, the interventions, materials and recruitment. Section 4 describes the findings. Section 5 discusses these findings, including some informal comparison between the arms of the study, and draws conclusions.

2 Method – overall study

2.1 Design

The trial had two arms. One arm tested the effectiveness of the novice driver interventions (mentoring agreements, telematics). The other tested the effectiveness of the learner driver interventions (logbook, hazard perception training, classroom-based education). See section 1.3 for high level descriptions of these interventions and section 3.3 for fuller descriptions. Figure 2-1 and Figure 2-2 show (again at a high level) the route participants took through the study. Participants were only able to sign up for one arm of the trial.

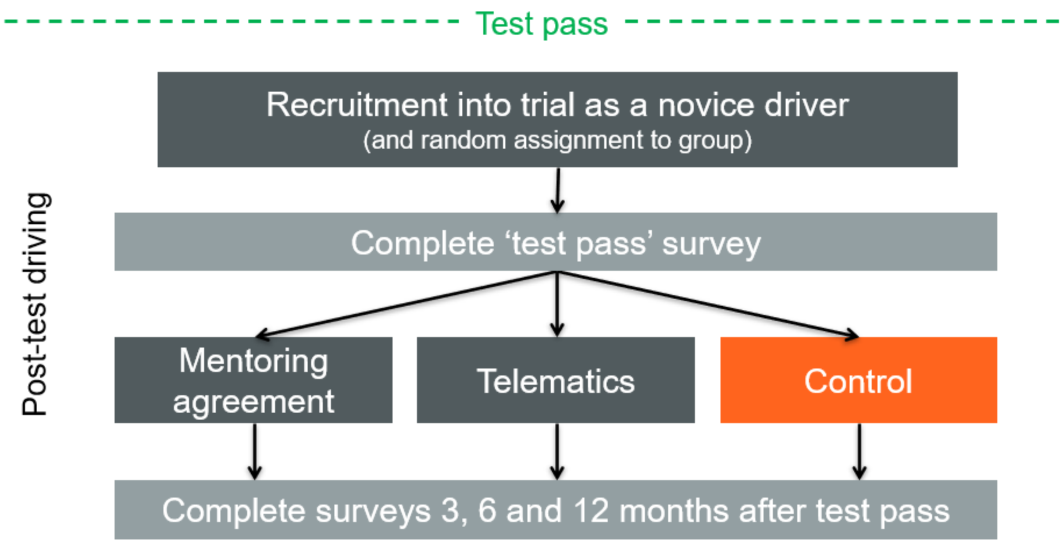


Figure 2-1: Design – novice driver arm

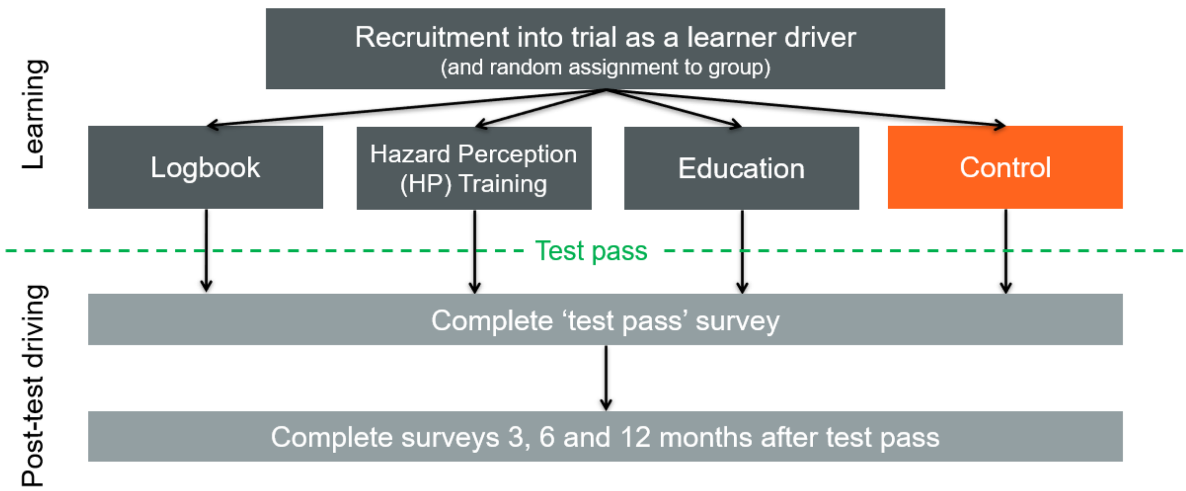


Figure 2-2: Design – learner driver arm

Because participants were assigned randomly to treatment or control groups in their respective arm of the study, the Driver2020 project is a type of randomised controlled trial (RCT). Incentives were required to reflect those that might be expected to be sustainable in any wider roll-out of a given intervention, meaning in practice that they were very modest.

The study could thus be better described as using a 'randomised encouragement' design (West et al., 2008). Unlike in classic RCTs in which participants are expected to adhere fully to a treatment protocol (and are given incentives to ensure this), in this design participants were randomly assigned only to an opportunity or 'encouragement' to receive an intervention. Crucially, the participants chose whether or not to engage with the intervention, and if they did engage, to what extent. This approach is useful when the incentives required to guarantee engagement would be unrealistic, undermining a generalisation of the findings to real-world roll-out. It is also useful when testing a voluntary (rather than legislatively-enforced) behaviour, as it allows exploration of engagement.

Thus, in simple terms Figure 2-1 shows that:

- Novice drivers were recruited into the trial shortly after passing their practical driving test, and then randomly assigned to either the mentoring agreement, telematics, or control group. They decided whether and how much to engage with any intervention they were offered during the next 12 months.
- They were immediately asked to complete their test pass survey on being recruited (because they had already passed their practical driving test).
- They were then asked to complete further surveys 3, 6 and 12 months after the date they passed their practical driving test.

Figure 2-2 shows that:

- Learner drivers were recruited into the trial and were randomly assigned to either the logbook, hazard perception training, classroom education, or control group. They decided whether and how much to engage with any intervention they were offered during their learning period (which varied in length, depending on the participant).
- They were asked to complete their test pass survey shortly after passing their practical driving test (if they passed within the timescales of the study).
- They were then asked to complete further surveys 3, 6 and 12 months after passing their practical driving test.

2.2 Approach to research questions

2.2.1 Research questions related to impact of the interventions

The main research question for each intervention was as follows:

- How effective is the intervention at reducing collisions in the first 12 months of post-test driving?

This was assessed by comparing the average number of self-reported collisions (controlling for other factors such as mileage) in the first 12 months of post-test driving in each intervention group with that of the corresponding control group.

A secondary research question for each intervention was:

- Does engaging with the intervention lead to other changes in relevant surrogate measures?

This was assessed by comparing the relevant surrogate measures for each intervention group (these differed by group and were largely self-reported) with the same in the corresponding control group. Surrogate measures included things like amount of pre-test practice (targeted by the logbook intervention) and hazard perception scores (targeted by the hazard perception training intervention).

In Great Britain it is not feasible to identify individual drivers in official casualty data (STATS19). This was the main reason why self-reported survey measures were used to collect data. The reliance on self-reported measures in answering surveys is also a cost-effective approach with such large samples – in total around 27,000 people registered into the study, though not all completed data collection. This approach may be subject to some measurement error and bias (for example people may respond in a socially-desirable way) and these limitations need to be taken into account when interpreting the findings; however, while not perfect, this approach has offered many insights in previous studies of this kind (e.g. Wells et al. 2008; Forsyth et al. 1995; Maycock et al. 1991).

2.2.2 Research question relating to engagement with the interventions

For all interventions, due to the fact that participants were free to choose to engage, the following research question was also asked:

- What were the factors that led people to engage with the intervention, or the barriers that stopped them engaging?

This was assessed through undertaking interviews with participants in each of the intervention groups who were known – usually through system data from each intervention provider – to have engaged to varying degrees (including some who had not engaged at all). Findings from these interviews – and from interviews with delivery partners – are reported in Hitchings et al. (2024).

The key focus of the engagement research was to complement the effectiveness evaluation, so that any interventions found to be effective could be rolled out with the best chance of appealing to or being encountered by young and novice drivers.

2.3 Timeline of the study

Figure 2-3 shows the relative timings of the recruitment and delivery of the learner and novice arms, including in relation to the COVID-19 pandemic.

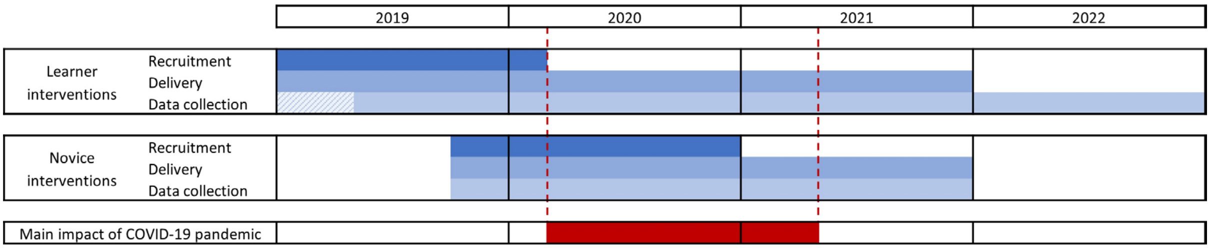


Figure 2-3: Timings for recruitment and delivery of the learner and novice interventions in relation to the COVID-19 pandemic

Recruitment for learners began in January 2019 and continued until early March 2020 (prior to the start of the COVID-19 pandemic). The delivery period for the learner interventions began at the same time and continued until December 2021, at which point delivery to those learner participants who had not yet passed their test was stopped. Data collection continued until December 2022 to allow 12 months post-test driving for those participants who passed their test close to this date.

Recruitment for novices began in October 2019 and continued until early January 2021. Delivery of the novice interventions began immediately and continued until December 2021 when the final novice participant completed 12 months post-test driving (and associated data collection).

Figure 2-3 also shows the timing of the main impact of the COVID-19 pandemic in relation to the recruitment and delivery of the interventions; this was from the beginning of the first lockdown in March 2020 until the end of the third lockdown in early May 2021. However, it should be noted that the impact of the pandemic, particularly on the waiting time for practical driving tests and therefore the pass rate of learner participants, continued beyond this point and was still ongoing at the end of the trial.

Details of the impact of the pandemic on the learner arm of the study and how these were addressed is discussed in the next section. The impact on the novice arm is discussed in Weekley et al. (2024a) which reports the results of the novice intervention trial.

2.4 The impact of the COVID-19 pandemic on the learner driver arm

2.4.1 Practical tests, theory tests and formal driving lessons (with Approved Driving Instructors (ADIs))

One of the key impacts of the pandemic on the learner arm was that practical driving tests were suspended (for all except key workers) during the periods of lockdown. These dates varied by country (and by region) and are summarised below in Table 2-1.

Table 2-1: Dates of practical driving test suspension in England, Scotland, and Wales during the COVID-19 pandemic

Practical (car) driving test	England	Scotland	Wales
1st lockdown			
Suspended	20 March 2020	20 March 2020	20 March 2020
Started	22 July 2020	14 September 2020	17 August 2020
2nd lockdown			
Suspended	5 November 2020	Not suspended	24 October 2020
Started	2 December 2020	Not suspended	9 November 2020
3rd lockdown			
Suspended	5 January 2021	26 December 2020	20 December 2020
Started	22 April 2021	6 May 2021	22 April 2021

Before and after the 2nd lockdown, all three countries were subject to localised restrictions known as tiers (England), protection levels (Scotland) and alert levels (Wales). In addition to the nationwide lockdowns listed above, practical driving tests were also suspended when an area was in Tier 4 / Protection Level 4 / Alert Level 4, which happened at various times.

Theory tests were also suspended (for all except key workers); the dates for these were broadly similar to the dates for practical tests.

Formal driving lessons were also cancelled over the same period. The advice for most driving school instructors and independent instructors was that lessons should not go ahead except for key workers who were preparing for an essential driving test.

The impact of these suspensions was significant for the learner participants (as it was for the participants in the novice arm – see Weekley et al. 2024a). With the exception of those who had already passed their test by March 2020 (15%), all were likely to have experienced an impact on their learning to drive process – by the cancellation of their tests (if booked), cancellation of lessons and the inability to book and take the test when ready. For learners in the hazard perception training group, it will most likely also have had an impact on access to the intervention as the modules were triggered by booking both the theory and practical tests.

As mentioned above, the impact of the pandemic continued after the suspensions of driving tests and driving tests were lifted; as a result of the restrictions the waiting times for practical tests was high compared with pre-pandemic. This means that a large number of learner participants may have been unable to take their test when desired; only 46% of the learner participants passed their test within the timeframes of the trial.

In March 2020, additional questions were added to the test pass survey to ask all participants who passed from that point on about the impact of the pandemic on their learning to drive process (for survey details, see the supplementary appendix document

(Weekley & Helman, 2024)). Questions were also added to the topic guides for the interviews in the qualitative evaluation, reported in Hitchings et al. (2024).

2.4.2 Reduced levels of driving

As well as formal suspensions, learner participants may have been affected by reduced levels of driving more generally. For example, during lockdowns, only essential journeys were permitted by restrictions.

For almost all participants this will likely have had an impact on the amount of driving experience they were able to get during this period (unless for example they were key workers) – either in their post-test driving, in their learning to drive process, or both. Only seven participants had completed 12 months' post-test driving by March 2020 and had therefore finished the trial before the pandemic started.

Impact on the learning-to-drive process has been identified through additional questions in the test pass survey as mentioned above. Impact on post-test driving has been estimated using self-reported measures of mileage during the survey reporting periods.

2.4.3 Cancellation of face-to-face classroom-based education courses

Before the pandemic, the education courses were delivered in a classroom setting. These were therefore cancelled due to lockdown restrictions. Unlike the other interventions, it did not seem likely that they could continue unchanged once driving tests resumed and therefore a decision was made to transfer the course to an eLearning delivery mechanism. The eLearning course was launched in October 2020; the changes in this intervention are discussed in section 3.3.3.

2.4.4 Implications

None of these impacts should introduce any bias into the study results as all groups (control and intervention) were subject to the same circumstances. Nonetheless the analysis is more complicated to account for changes at different times for the participants (in exposure to driving, for example) and this is discussed further in later sections of this report and is also addressed in Weekley et al. (2024a) which covers the novice driver interventions.

Another implication of the impacts was on the original intention in the study to achieve an approximate alignment in time of the post-test driving periods in the learner and novice arms. COVID-19 disrupted recruitment (and test-passing) substantially, and this means that the alignment was much less than anticipated.

The generalisability of the findings to future circumstances (where there is no pandemic) may be affected; the pandemic will have significantly altered the learning to drive process and post-test driving of most of our trial participants. For example, the time to obtain a licence for this cohort is likely to have been significantly increased compared with what has been observed before. This provides both challenges for the current study and an opportunity to gain further insight into the effect of age and experience on young drivers' collision risk.

3 Method – learner driver arm

3.1 Participants and recruitment

The eligibility criteria for learners were that they were aged between 17 and 24 years, and either intended to begin learning to drive in the next two months or had already begun learning to drive but with only up to 10 hours on-road practice so far. In reality, participants were able to register with the trial before the age of 17 provided that they would turn 17 and begin learning to drive within the next two months. They were also required to have or be planning to have some access to a vehicle for the 12 months after they pass their test. The various recruitment methods directed participants to the registration survey, which assessed whether participants met these criteria .

The target sample size was 2,036 per group (this achieving 80% power to detect a 20% reduction in collision involvement from an expected baseline of 15%; so a change from 15% of drivers having a collision within their first 12 months of driving, to 12%).

The main recruitment mechanism was a leaflet included with the provisional licence sent by the Driver and Vehicle Licensing Agency (DVLA). A leaflet was included with all provisional licences issued from 5th April 2019 until approximately 25th February 2020. Figure 3-1 shows the double-sided leaflet insert; note that there was also a bilingual version for use in Wales.

Figure 3-1: Recruitment leaflet included with provisional licences.

A secondary recruitment method used was advertising through approved driving instructors (ADIs); ADIs were given digital versions of the leaflet to show to learners.

Participants were asked in the registration survey how they had heard about the study; 94% of registrants indicated that they heard about the study from the leaflet; of the remaining 6%, 3% stated that they had heard about the study through an announcement on social media, 1% from a friend and 2% from an alternative mechanism, for example driving instructor or other. The learner recruitment period officially opened on 9th January 2019 with the announcement of the trial on social media and the launch of the project website, and with ADIs being asked to advertise; however the leaflet recruitment mechanism was not in operation until 5th April 2019 as mentioned above; Figure 3-2 shows the number of participants registered per month, clearly showing the impact of the primary recruitment mechanism.

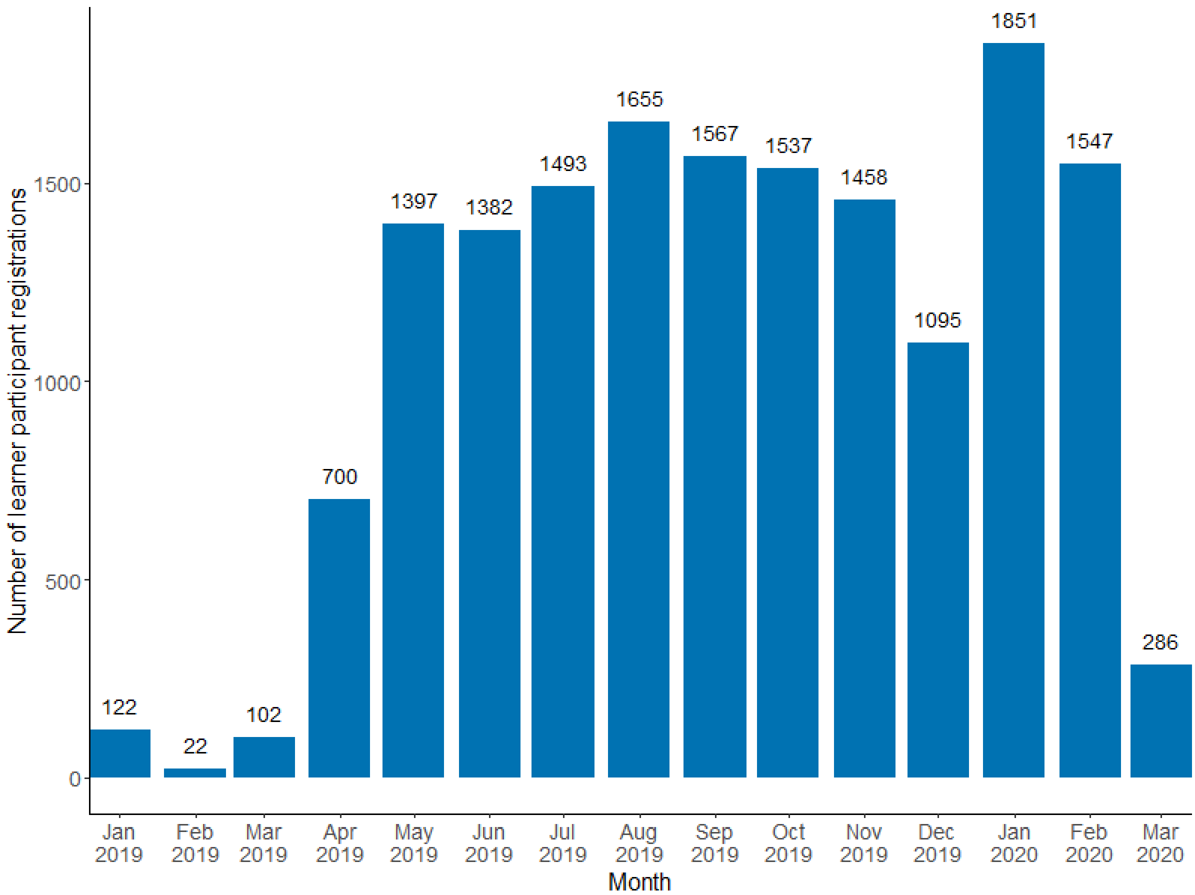


Figure 3-2: Number of learner participant registrations per month during the recruitment period (Jan 2019 to Mar 2020)

Note that recruitment ended on 8th March 2020, as target recruitment numbers for each group (approximately 4,000 for each group) had been achieved.

The total number of learner participants recruited into the learner driver arm of the study was 16,214 across Great Britain. The gender distribution and mean age for each group of recruited participants is shown in Table 3-1

Table 3-1: Mean age and gender distribution for the total number of learner participants recruited

Group	Total recruited	Female	Male	Mean age
Logbook	4438	60%	40%	18.18
Hazard perception training	3851	60%	40%	18.20
Education	3850	59%	41%	18.24
Control	4075	61%	39%	18.18
Total	16,214	60%	40%	18.20

Please note that this is all participants recruited, not those who provided data through surveys. For demographics of participants who provided data, and detail on those who were included in the final trial sample for analysis, see Section 4.

Participants were provided with a £5 retail voucher (or equivalent donation to charity) for each of the four data collection surveys they completed (detail on the surveys is included in section 3.3.4). In addition, all those who completed the 12-month survey were automatically entered into a prize draw to win one of four top prizes (a year's free car insurance) or other prizes such as iPads and retail vouchers. These incentives were designed to encourage completion of the data collection surveys and minimise drop out of the participants from engaging with this element of the trial. The incentive structure was informed by previous studies including Helman et al. (2017). In line with UK law on prize draws, people were also informed that they could send their contact details to be entered into the prize draw, even if they did not take part in the trial data collection.

3.2 Design

3.2.1 Logic models

The logic models created for the logbook, hazard perception training and education interventions are available in the supplementary appendix document (Weekley & Helman, 2024). A logic model is a diagram that shows how an intervention is expected to produce its effects – how it is meant to work (O'Cathain et al., 2019; Smith, Li & Rafferty, 2020).

3.2.2 Group assignment

The design of the study was a randomised encouragement design, meaning that participants in the intervention groups were given the opportunity to engage with their intervention, but were not required to do so. Registered learner participants were randomly assigned to the logbook, hazard perception training, education or control group. The assignment was weighted slightly more heavily towards the logbook group than towards the other groups, as it was judged that the drop-out rate from this intervention may be higher (although this did not turn out to be the case). The randomisation therefore proceeded as follows:

- The database assigned a given participant to either the logbook, hazard perception training, education, or control group by assigning a study ID number randomly

(weighted according to how many 'vacancies' – compared with the target number – in each group remained)

- For every 2,850 control group assignments, 2700 assignments were made to the hazard perception training group and the education group, and 3,100 to the logbook group.

Appendix B shows the data flows and way in which participants were handled when brought into the study (see also section 3.4).

3.2.3 Analyses

Two analyses were undertaken. First, an 'intention-to-treat' analysis compared the number of self-reported collisions in the first 12 months of post-test driving between the four groups, including every participant for whom there was survey data (even if they did not engage with the intervention to which they had been given access). The purpose of this approach is to understand the real-world effectiveness of an intervention, including any effects of non-engagement (Montori & Guyatt, 2001). Generalised linear modelling was used in this analysis. First, a base model was constructed to understand the role of exposure, experience, age, and gender on collisions. This base model was then used to check for the effect of the COVID-19 pandemic, and the interventions. See section 4.4 for a detailed description of the modelling approach.

The two-person team of statisticians undertaking the intention-to-treat analysis was blinded to group identity until after the full analysis had been run. Any variables which might give away group membership (for example the question in the 12-month survey on which interventions the participant had heard of) were removed at this stage of the analysis.

The second analysis was a 'per-protocol' analysis looking only at the impact of the intervention on those people who were known to engage to some degree. The purpose of such an analysis is to examine the potential efficacy of the interventions in individuals who choose to engage in the event that the numbers of people who engaged were low. In this analysis some of the non-collision measures in the surveys were used to assess efficacy; some of these were general surrogate measures for safety, and some were those that matched the intended outcomes of the interventions.

The following were used as general surrogate measures for safety:

- Self-reported near misses
- Self-reported Driving Events scale score (items on hazard involvement)
- Self-reported frequency of exceeding the speed limit

For the logbook intervention per-protocol analysis the following additional surrogate measures were compared between the logbook and control groups:

- Self-reported number of hours of on-road practice in the learning phase
- Self-reported proportion of road types and driving types experienced in the learning phase

For the hazard perception training intervention, the following additional surrogate measures were used:

- Number of attempts required to pass the theory test
- Self-reported driving style

Finally for the education intervention the following additional surrogate measures were used:

- Self-reported number of hours of on-road practice in the learning phase
- Self-reported proportion of road types and driving types experienced in the learning phase
- Self-reported driving style
- Self-reported proportion of mileage in the first 12 months post-test done at night, with passengers, or when tired
- Self-reported attitudes towards speed enforcement and new driver restrictions.

It should be noted that surrogate measures were used for the ‘per-protocol’ analysis instead of collisions, as the samples of engagers were far too small to permit a reliable analysis of the collision measure. The sample sizes for engagers were sufficient to support analysis of the surrogate measures, as these do not suffer from the very low baselines/high variability associated with the collision variable (which consequently requires much greater sample sizes for analysis).

Prior to these main analyses, that were focused on the research questions, other comparisons were made between groups to ensure that they were matched on variables such as age, gender and personality variables; this was especially relevant for the per-protocol analysis to check that those who engaged with the interventions were matched with appropriate participants in the control group.

Statistical tests were used throughout this study to test for differences between groups (for instance, control, logbook, hazard perception training or education course groups, or between pre-pandemic and post-pandemic groups).

In all cases where statistical tests were used to compare data, the convention from the behavioural sciences of reporting *p*-values was adopted. The probability value, *p*-value, is used to determine statistical significance. However, statistical significance alone can be misleading in trials with very large samples because it is influenced by the sample size. Therefore, this study presents the effect sizes where the results were statistically significant. Effect sizes help understand the magnitude of differences found. Both are essential to understand the full impact of the interventions. Findings are statistically significant only when explicitly stated.

The meaning of an effect size varies on the type of statistical test that is being used; the same effect size value has different meanings for different statistical tests. To give an indication of real-world impact, values of a given effect size statistic are described as ‘negligible’, ‘very small’, ‘small’, ‘medium’ or ‘large’. For more information on statistical significance, effect sizes and details on the specific statistical tests used for each comparison, see Appendix C.

Throughout the analyses, appropriate graphical outputs have been used to present the data and give a quick visual representation of the distribution of values. This includes tables and charts such as histograms, bar and line plots, boxplots and violin plots. Histograms are used to show the frequency of data in particular groups. (Please note that, due to rounding, the sum of the data labels on some histograms does not equal exactly 100%.) Box plots are a visual representation of the inter-quartile ranges (25^{th} and 75^{th} percentiles) and median (50^{th} percentile or 'average') for a continuous variable (values outside of the 95% confidence interval are shown as individual data points). Violin plots are similar to boxplots but also show the density plot (where the width of each curve corresponds with the frequency of data points in each region) of a continuous variable.

3.3 Materials

3.3.1 Intervention: Logbook

The logbook intervention (delivered by the Driving Instructors Association - DIA) was a smartphone app. Participants allocated to this group were invited to download the app and set up a user profile so that they could use it to help guide their learning to drive.

The app itself provided advice to participants on how much and what types of practice to try and attain throughout their learning to drive period. This included:

1. Trying to achieve at least 100 hours of on-road practice (combined between ADI lessons and private practice) if possible
2. Trying to make sure the following situations were covered in their on-road practice if possible:
    - Driving at night
    - Bad weather
    - Country roads/lanes
    - Congested busy town centres
    - Motorways
    - Dual carriageways
    - Residential areas
    - Driving with passengers
    - Using a satnav
    - Using road signs

Users could log their individual lessons and private practice sessions, applying the above categories to help track progress (see Figure 3-3). People were able to check within the app to check how they were doing in terms of covering the different driving situations – a running total of driving that had been logged was kept.

The people in the logbook intervention group were also given the opportunity to sign up with a learner driver insurance product with a specialist provider (https://www.collingwood.co.uk/learner-driver-insurance/) at a small discount (£25 discount for a 6 month learner policy or £50 discount for an annual learner policy, and £50 discount for a young driver policy). This offer was available to any participant in this group who downloaded the app. The aim was to mitigate against the insurance cost that can impede learners getting more on-road practice.

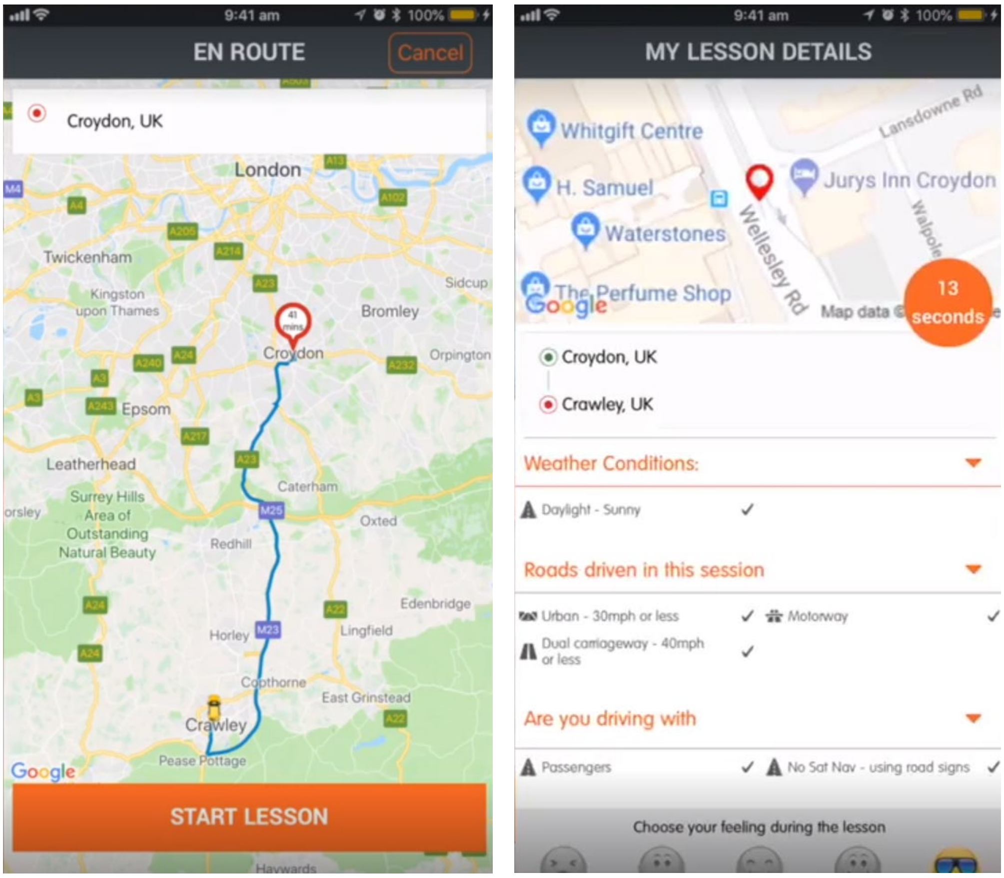


Figure 3-3: Screenshots of logbook app

3.3.2 Intervention: Hazard perception training

The hazard perception training intervention (delivered by DriverMetrics) used an online eLearning platform to deliver three modules (each approximately 35 minutes in length) of training and testing clips designed to increase hazard perception skill in drivers. Participants assigned to this intervention group were asked to sign up with the system run by DriverMetrics, and were then invited to do the three modules at the following time points:

- Module 1 – on booking their theory test: The reasoning for this timepoint is that learners should be motivated to pass the hazard perception test element of their theory test.
- Module 2 – on booking their practical test: The reasoning here is that the skill could be boosted close to a point at which solo driving might soon begin.

- Module 3 – on passing their practical test: The reasoning here is that the skill is boosted in early driving, when risk is high.

A range of hazard categories were included (weather, nighttime, bends, cyclists and pedestrians, motorcyclists, motorways, moving off, stopping); some other topics were covered due to their inclusion in the driving test (for example, using satnav) to increase face validity – the perceived relevance of the training to the basics of the driving task, or as advanced concepts (prioritising hazards, safety margins, looking for clues). Figure 3-4 shows some example screenshots.

Figure 3-4: Example screenshots of hazard perception clips from the intervention

3.3.3 Intervention: Education course

The education intervention (delivered by Agilysis) was originally designed as a classroom-based one-day (six hour) course (although note that these courses were delivered outside of formal education locations).

The design was based around interactivity and discussion, with trained, credible facilitators.

The course was structured into five sessions, designed to be taken at any point in the learning to drive process:

- Encouragement of self-reflection of the learners' own personal goals for driving and the identification of driving challenges (both vehicle manoeuvring and situational awareness).
- Hazard perception and hazard prediction (including active interaction with video-based stimuli, and theoretical consideration through questions and answers).
- Insight into their own limitations ('how to drive'), including speed choice and close following.
- Driving choices and planning for different scenarios ('when to drive'), including coping options and strategies.
- Specific personal goal setting in relation to skills practice and positive behaviour planning.

During the COVID-19 pandemic, it was decided that it would not be possible to continue with the classroom-based education course even after restrictions were lifted. Therefore, the intervention was redesigned for an online delivery mechanism.

The structure of this was five modules, in line with the five sessions above, which were available for completion whenever the participant wanted to engage. They were accessible on laptops, tablets, or smartphones. Facilitators were still used through recorded content and voiceovers to the modules.

The content was approximately the same length as the classroom session, although this could vary with how participants interacted with the modules, and the modules did not have to be completed in one sitting. All modules and activities in both versions were based on a combination of behaviour change techniques. Once either version of the course was completed, participants were offered access to a follow-up app that they could download; this provided support resources and functionality to review and record their activities in line with their goal setting.

3.3.4 Surveys

Five surveys were used to collect data from all participants, in all intervention and control groups. These are described below and are included in full in the supplementary appendix for the project (Weekley & Helman, 2024).

3.3.4.1 Registration survey

The registration survey (10-15 minutes, online) first collected information about participants to check they were eligible to register with the study (see section 3.1 for criteria).

For those participants who were eligible, the remainder of the survey served as an informed consent form including relevant data protection information, and collected contact details, driver licence number, information on education and employment, postcode, how the participant had heard about the study, and finally a 30-item form of the Big-Five Personality Inventory (Soto & John, 2017) to be used as a matching variable (this was one of a number of measures that was used to control for potential bias in engagers, relative to control participants – see section 4.5.2).

3.3.4.2 Test pass survey

The test pass survey first collected name and participant ID for the purposes of data matching. The remainder of the survey asked a range of questions about the participant's experience learning to drive (for example the types of roads they practised on, the amount of practice they had with instructors and with other supervising drivers), various items on attitudes to post-test restrictions and more general enforcement of road laws, information on education and employment, and items around access to a vehicle.

3.3.4.3 3-month, 6-month and 12-month post-test surveys

The 3-month, 6-month and 12-month post-test surveys began with questions collecting name and participant ID (for the purposes of data matching). The remainder of the survey asked a range of questions about the participant's experience driving post-test. The main items in the 3-, 6- and 12-month post-test surveys were the same; the difference in the surveys was that they asked about different periods of time (the first three months post-test, the fourth to sixth months, and the seventh to twelfth months respectively). The items included were:

- The number of collisions in which people had been involved during the period of interest, if any (and some further questions about when, where, and how each collision occurred)
- Mileage driven
- Proportion of mileage driven on different road types
- Proportion of mileage driven with a peer-age passenger, driven in the dark, driven on wet roads, and driven for work
- Near misses
- Six of the seven items from the Hazard Involvement/Driving Events scale from Quimby et al. (1999)
- Frequency of driving while tired, and driving over the speed limit
- Any limits or restrictions on their driving during the period of interest, for example less access to a car than expected, parents imposing restrictions, impacts of the COVID-19 pandemic or other changes in circumstances.
- The driving style scales from Guppy, Wilson and Perry (1990)
- Confidence in driving ability
- Attitudes towards enforcement of various road safety laws, and restrictions and limits for learner and new drivers
- Access to a vehicle and details of insurance policy type
- Levels of education and employment details

A final question in the 12-month survey asked participants if they had come into contact with any of the other interventions (a check for spill-over effects – for example if someone had come into contact with one of the interventions through contact with a friend in the study – see section 4.3.3).

3.4 Procedure

This section describes the procedure for each participant on entry into the study and the subsequent communication that took place. This procedure and the related data flows are illustrated in Appendix B.

Each participant who completed the registration survey and was eligible for the study was assigned to a group on the basis of the process described in section 3.2.2 and sent a welcome email. Those in one of the intervention groups were told that the relevant intervention provider would be in touch (for the logbook, DIA; for the hazard perception training, DriverMetrics; for the education course, Agilysis). Those in the control group also received the welcome email which confirmed that they would not be offered an intervention, but would still receive invitations to complete the surveys once they passed their test.

All learner participants received keeping-in-touch emails every six months until they had passed their test (although note that these were put on hold during the lockdowns related to the COVID-19 pandemic).

Once the welcome email had been sent, the details of those assigned to one of the intervention groups were passed on to the relevant intervention provider. After they received the contact details of the participants in their specific intervention group, providers contacted them to invite them to register with the specific intervention. With the exception of the keeping-in-touch emails, all communication with learner participants before they passed their test was via the intervention provider. However, all communications were branded as Driver2020 to minimise confusion.

For DIA (logbook), participants were sent a welcome email with instructions for downloading the app and how to create an account, as well as access to the website for more information. This email also included how to nominate a parent/guardian or driving instructor to also take part. If a parent/guardian or driving instructor were nominated, they also received a welcome email. During delivery, each time a learner recorded a session on the app, the parent/guardian or instructor would be prompted by email to review the session, and in turn the learner emailed to view the feedback. Following the lifting of COVID-19 restrictions, reminder emails were sent to those learners who had created an account to encourage continued use of the app.

For DriverMetrics (hazard perception training), participants were sent a welcome email to explain that they would be given access to each module at the associated trigger points. Once it was confirmed that a participant had booked their theory test (see below), they would receive an email asking them to activate module 1; there was also a follow-up email reminder if they did not activate it. Participants were then emailed again to activate modules 2 and 3 when they booked and passed their practical test respectively.

For Agilysis (education course) the communications were slightly different for the classroom-based format and the online delivery. All participants received a welcome email shortly after they registered to invite them to attend a classroom-based course and with instructions for booking. Those who booked a course were emailed follow-up information immediately and texted with a reminder one week before the course and again the day before the course. During the delivery period for classroom-based education, some targeted invitations were sent to all participants in this group who had not yet attended a course and were within a 30-mile radius of the venue location.

When the online delivery mechanism was launched, all participants who had not completed a classroom course were sent an invitation with instructions for how to access the modules. This invitation was then repeated every month to those who had not yet begun a module. When at least one module had been started, participants received a reminder once every two weeks unless all modules were completed.

Participant details for all current learners in the trial were passed to the Driver and Vehicle Standards Agency (DVSA) twice a week to confirm whether they had booked or passed a practical driving test (and for those in the hazard perception training group, whether they had booked or passed a theory test). Once it was confirmed through this process that a participant had passed their practical test, all subsequent communication was with TRL regarding the data collection surveys; the relevant intervention providers were informed and the participants were no longer contacted in relation to the interventions.

Firstly, participants were invited to compete the test pass survey. All participants then were sent emails by TRL at 3, 6 and 12 months after their test pass date providing links to the relevant data collection surveys. If a survey was not completed, the participant received an email reminder with the survey link every week, up to a maximum of six reminders. Text message reminders (also including the survey link) were sent at the same time as the third and fourth emails. These reminders were also used for the test pass survey. For some participants who had not responded to reminders for the 12-month survey, the TRL research team also telephoned them to encourage them to complete the survey.

3.5 Limitations

There are a number of limitations of the research that need to be considered when drawing conclusions. These are mentioned throughout the report and are listed here for clarity.

1. The data were gathered during the COVID-19 pandemic, which would have affected learning and post-test driving. Generalisability to future circumstances in which there is no global pandemic may be affected.
2. The sample is biased towards females and is an opportunity sample. Therefore, any generalisation of the findings to the population of interest (novice drivers aged between 17 and 24 years of age in Great Britain) needs to be done with caution.
3. The self-selecting nature of the sample (both in terms of registering for the trial and in terms of providing survey data) means that the findings cannot necessarily be generalised to all learner drivers in Great Britain. However, the design (randomised allocation to groups) does mean that self-selection bias should not play any role in group differences.

4. The self-reported nature of most of the data means that conclusions again require caution; it is possible that the reported data are biased to some degree with social desirability; although this is not critical for the main comparisons between groups, it may mean that reported levels of behaviour are different to what would actually be observed. Self-reported data also reflect the interpretation of participants to survey questions.
5. Given the very low numbers of participants engaging with the interventions, and the low ‘dosages’ or ‘amounts’ of contact with interventions in those participants who did engage, the findings on collision reduction reflect only one potential set of roll-out conditions for the interventions.
6. Some groups had smaller-than-planned samples, meaning that some very small effects may have been missed due to lower statistical power.
7. The findings related to surrogate measures provide us with evidence of the behaviour change that results from the interventions as these analyses only include those who engaged; however, the very low numbers engaging with some interventions mean that some very small effects may have been missed.
8. The findings only cover the first year of driving and cannot be extended beyond this period (for example we cannot say whether behaviour would be impacted beyond this period, or whether any changes seen would persist).

4 Results – learner driver arm

4.1 Introduction

This section presents the results of the learner driver arm of the Driver2020 study. The focus is on the intention-to-treat analysis – sections 4.2 to 4.4 cover the trial sample used, exploratory analysis and statistical modelling. Section 4.5 presents the per-protocol analysis.

4.2 Trial sample

The final dataset to be used for the learner driver 'intention-to-treat' analysis was collated on 30th January 2023. Any survey responses received after 23:59 on 30th January 2023 have not been included in analysis.

Table 4-1 shows the completion rate for each of the four data collection surveys for learner participants, along with the total number recruited, and the total who passed their test, for each group. The percentage given for each survey is the percentage of participants who completed the survey from those that qualified for it; not all participants who passed their test reached all three subsequent time points due to the overall timescales of the trial. In each case, the total that qualified for the survey is given in brackets after the total that completed the survey.

Table 4-1: Number and percentage of learner participants who completed each survey (Rec: the total number recruited, Pass: the number who passed their test)

	Logbook (Rec: 4,438, Pass: 2,403)		HP Training (Rec: 3,851, Pass: 2,063)		Education (Rec: 3,850, Pass: 2,024)		Control (Rec: 4,075, Pass: 2,235)		Total (Rec: 16,214, Pass: 8,716)	
Test Pass	1,583 (2,400)	***66%***	1,400 (2,059)	***68%***	1,243 (2,024)	***61%***	1,555 (2,233)	***70%***	5,781 (8,716)	***66%***
3-month	1,341 (2,322)	***58%***	1,122 (1,990)	***56%***	995 (1,976)	***50%***	1,315 (2,177)	***60%***	4,773 (8,465)	***56%***
6-month	1,200 (2,220)	***54%***	1,012 (1,908)	***53%***	910 (1,896)	***48%***	1,215 (2,074)	***59%***	4,337 (8,098)	***54%***
12-month	1,061 (1,967)	***54%***	884 (1,692)	***52%***	796 (1,686)	***47%***	1,037 (1,829)	***57%***	3,778 (7,174)	***53%***

Note that, for each survey, the number and percentage shown represents the completion for that survey point, independent of whether previous surveys had been completed; not all the participants who completed the 12-month survey had also completed all previous ones.

Table 4-2 shows the number of participants in each group who completed all four surveys and that number as a percentage of those who were eligible to complete all four surveys (for learners, this was participants who passed their test and completed 12 months post-test driving).

Table 4-2: Number of participants in each group who completed all four surveys.

Group	Number of participants with all four surveys completed (as % of total eligible)
Logbook	916 (47%)
Hazard Perception Training	769 (45%)
Education	682 (40%
Control	925 (51%)
Total	3,292 (46%)

The original sample size calculations for the trial resulted in a target of 2,036 participants per group; due to the impact of the COVID-19 pandemic on the overall timescales of the trial these targets were not achieved for the learner arm. This is likely to be mainly due to the fact that around half of the sample of learners did not pass their test in time to provide data on post-test driving.

In addition, there were 115 participants who completed the test pass survey and also provided data by phone call at the 12-month time point. These participants did not fill out the 12-month survey but provided estimated mileage and collision data (number of collisions) and near misses for the entire 12-month period. Since these participants provided the required data for modelling the number of collisions in the first 12 months after passing the test, they were included in the trial sample for the intention to treat analysis. This sample – consisting of these additional participants and those that completed all four surveys – is referred to as the 'trial sample' and is shown in Table 4-3.

Table 4-3: Number of participants in each group for the trial sample

Group	Trial sample
Logbook	944
Hazard Perception Training	803
Education	712
Control	948
Total	3,407

4.3 Exploratory analysis

This section presents the exploratory analysis of the trial sample, their learning to drive experience, and their experiences during their first year of driving. The aim of this was to explore whether characteristics differed between the four groups: control, logbook intervention, hazard perception training intervention, and education course intervention, to assess the comparability of the groups for further analysis. To compare some of the characteristics, answers to all the surveys are required and therefore the participants supplying phone call data were not included, only those who had completed all four surveys. Since the former group was very small (around 30 per intervention), not including them is unlikely to impact any of the comparisons. Where these additional participants were not included in the analysis, this is indicated by the base sizes provided with the charts.

The demographics of the trial sample are explored in section 4.3.1, and these are then compared with the demographics of test passers nationally in section 4.3.2 to examine how representative the trial sample is of the national population. This is for information only, as the study was always targeted at younger learner drivers than the national population of test passers. In addition, in studies of this kind females are typically over-represented (for example see Wells et al., 2008; Helman et al., 2017) as was the case in this study.

Section 4.3.4 presents analysis of the conditions in which participants in each group learned to drive and drove during their first post-test year, such as different road types, weather, and light conditions. The impact of the COVID-19 pandemic on participants and their learning to drive and post-test driving experiences is explored in section 4.3.5.

A summary of the number of collisions reported by participants is presented in section 4.3.6, and compared between different demographics, post-test mileage levels, and intervention groups. Finally, a brief summary of the attrition analysis – carried out to determine if there were differences between the final sample and those recruited – is described in section 4.3.7 .

The key insights from the exploratory analysis of the learner driver sample are as follows:

- The sample was biased towards females (64% of the sample was female) meaning caution is needed in generalising findings.
- Around 84% of the sample was in full time education when registering to take part – this was higher than in the novice arm, as expected, due to learners registering at a younger age (when they began learning, rather than after they had passed their test).
- There was no evidence of spill-over or contamination effects (people in one group being exposed to interventions from the others).
- Participants who passed their test after the COVID-19 pandemic reported a number of impacts from the pandemic on their learning to drive, mainly tests being postponed or cancelled, not being able to book practical tests as soon as they would like, and having less practice with a driving instructor.
- Participants who passed their test before the COVID-19 pandemic reported lower annual mileage than those who passed after.
- There were no significant differences between the groups for any of the characteristics explored, meaning that the groups would be suitable for comparison in the intention-to-treat analysis.
- There were changes to the characteristics of the sample over time as some types of people were more likely to drop out of the study than others, but these changes were consistent between the groups.

4.3.1 Demographics

This section presents a summary of the demographics of the trial sample. The distributions of age, gender, Social Deprivation Index (SDI), and education status were compared between the four groups (control, logbook, hazard perception training, and education course) but statistical tests found no differences between groups.

The participants included in the analysis presented in this section are those who answered all surveys or provided 12-month data by phone (3,407 participants).

4.3.1.1 Age

Figure 4-1 shows the proportion of participants in each group by age at registration. The majority of trial participants were 16 or 17 years old when they registered for the trial. These two age categories account for 9% and 67% of all trial participants respectively. Differences in age distribution between groups were not statistically significant. The mean age (at registration) of a participant in the trial was 17.8. This is over a year younger than in the novice arm, which is due to the fact that learners were recruited when they began learning, while novices were recruited when they had passed their practical test. Also, it should be noted that learners needed to be between 17 and 24 years old when they started learning to drive.

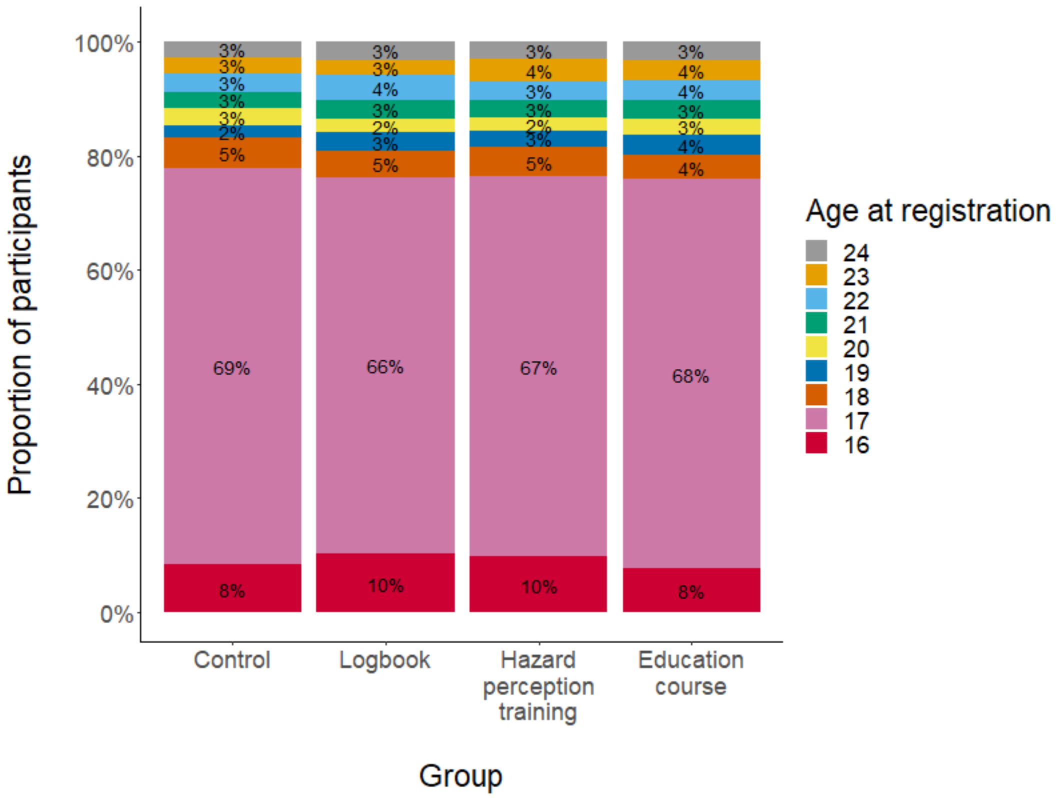


Figure 4-1: Proportion of trial participants in each group by age at registration (N=3,407)

Figure 4-2 shows the proportion of participants in each group by age at test pass.

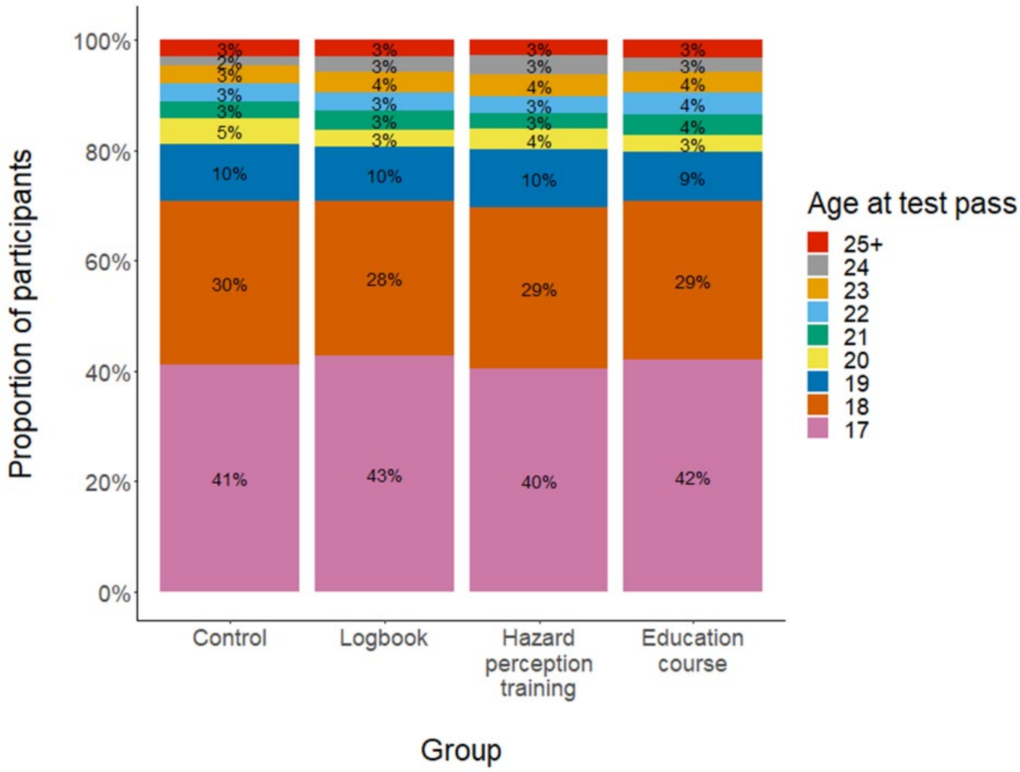


Figure 4-2: Proportion of trial participants in each group by age at test pass (N=3,407)

4.3.1.2 Gender

The proportion of male and female participants in each group is shown in Figure 4-3. The trial sample was nearly two-thirds female (64%); as noted previously, in studies of this kind females are typically over-represented (for example see Wells et al., 2008; Helman et al., 2017). This means that findings cannot be generalised to the population of young and novice drivers in Great Britain without some caution. There were small differences in gender distribution between groups, however these were not statistically significant.

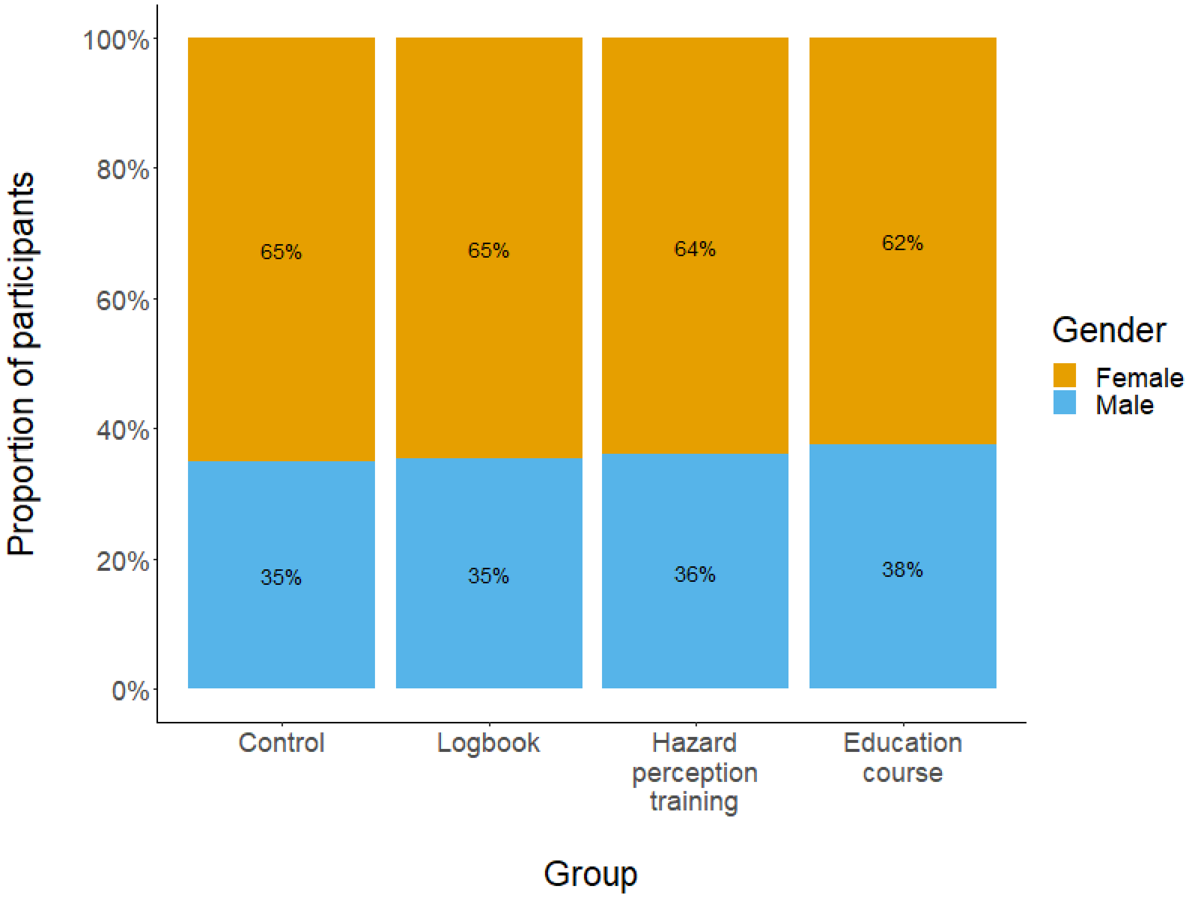


Figure 4-3: Proportion of trial participants in each group by gender (N=3,407)

4.3.1.3 Social Deprivation Index

Index of Multiple Deprivation (or social deprivation index) is the official measure of relative deprivation for small areas (or neighbourhoods) in the UK. Data from England, Wales, and Scotland (gov.uk, 2019; gov.scot, 2020; gov.wales, 2019) were downloaded and combined to create one dataset. This was matched to the participant data by location information (postcode at time of registration) and the social deprivation ranks were then assigned to four quartiles from most deprived to least deprived.

Figure 4-4 shows the proportion of participants in each group by Social Deprivation Index quartile. There are small (not statistically significant) differences between groups, particularly in the Q4 (least deprived) proportions, however a similar proportion (just over 60%) are in Q3 and Q4 combined for all groups. Q1 (most deprived) accounted for the smallest proportion of participants across all groups (17%), while 21% and 26% of participants were in Q2 and Q3 respectively, and the largest proportion (36%) was in Q4 (least deprived).

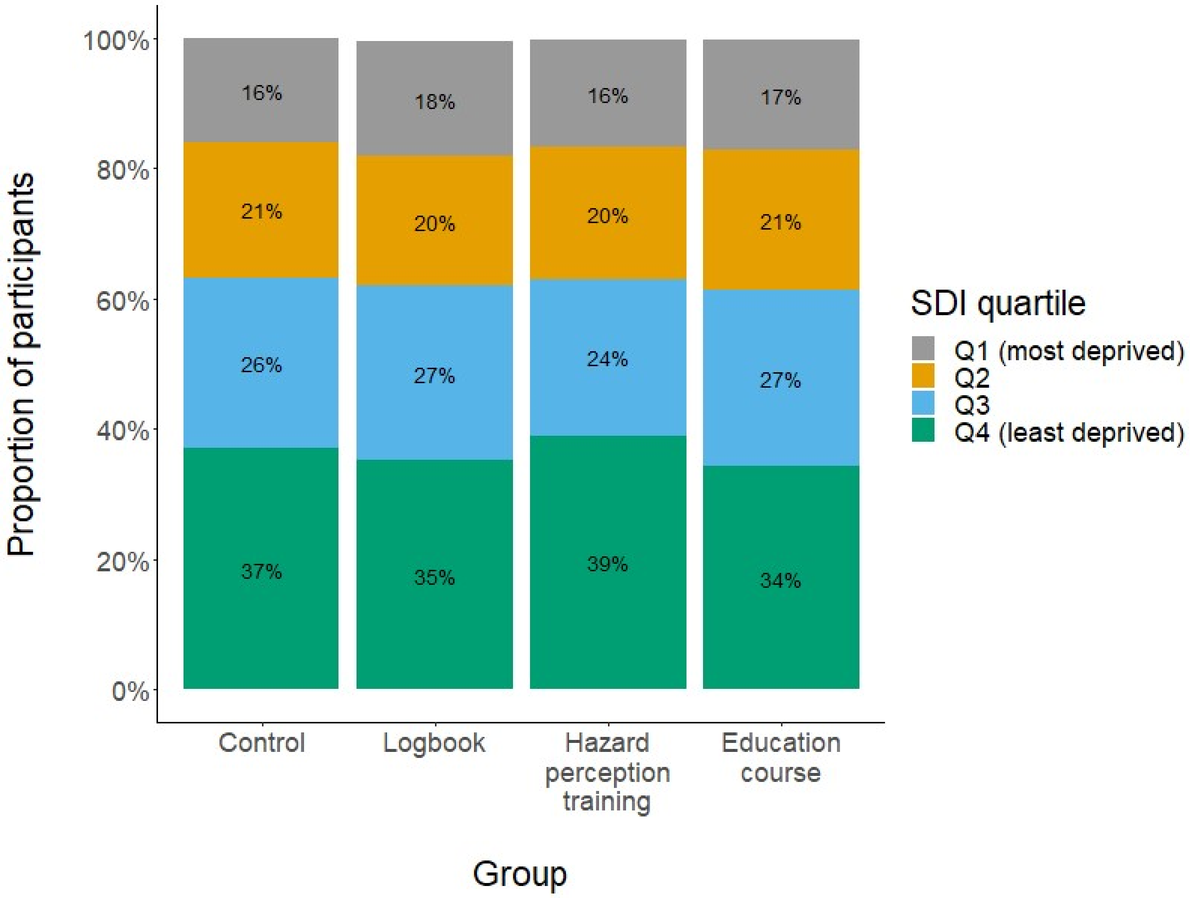


Figure 4-4: Proportion of trial participants in each group by Social Deprivation Index quartile (N=3,407)

4.3.1.4 Education status

Figure 4-5 shows the proportion of participants in each group who were in full-time education when they registered for the trial. The majority of trial participants in the learner arm (84%) were in full-time education when they registered. This proportion was similar between groups, albeit with the control group having a slightly higher proportion (86%). These differences were not statistically significant.

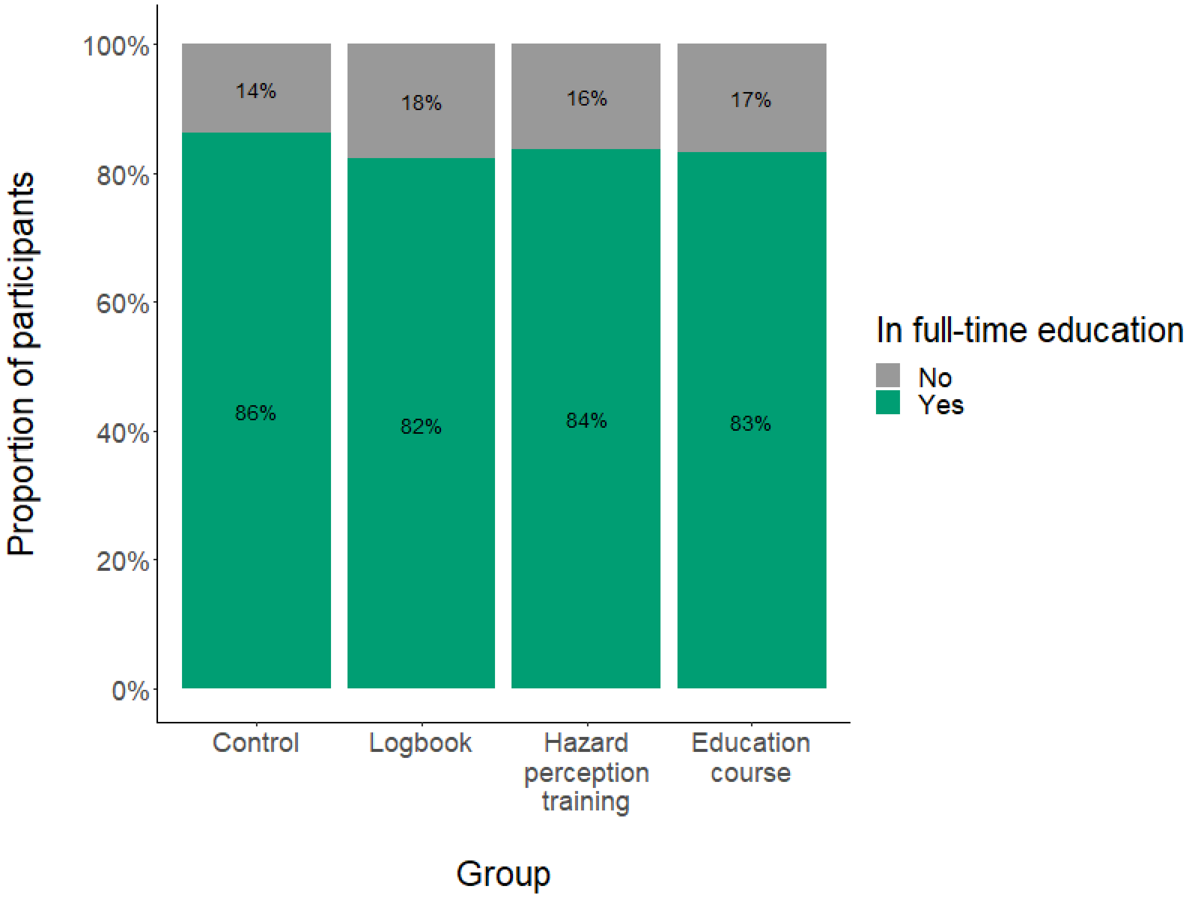


Figure 4-5: Proportion of trial participants in each group by education status (N=3,407)

4.3.2 Comparison with other data sources

The age and gender distributions of the trial participants were compared with test passers in the whole of Great Britain during the period April 2019 to March 2022 as published by DfT in table DRT0203 (DfT, 2022). These comparisons were done to explore how similar those registered in the trial were to the population of test passers nationally.

The results of these comparisons are presented in Figure 4-6 (age) and Figure 4-7 (gender), and showed that the trial sample includes higher proportions of younger people and females than seen across all test passers in Great Britain during a similar period.

The participants included in the analysis presented in this section are those who answered all surveys or provided 12-month data by phone (3,407 participants).

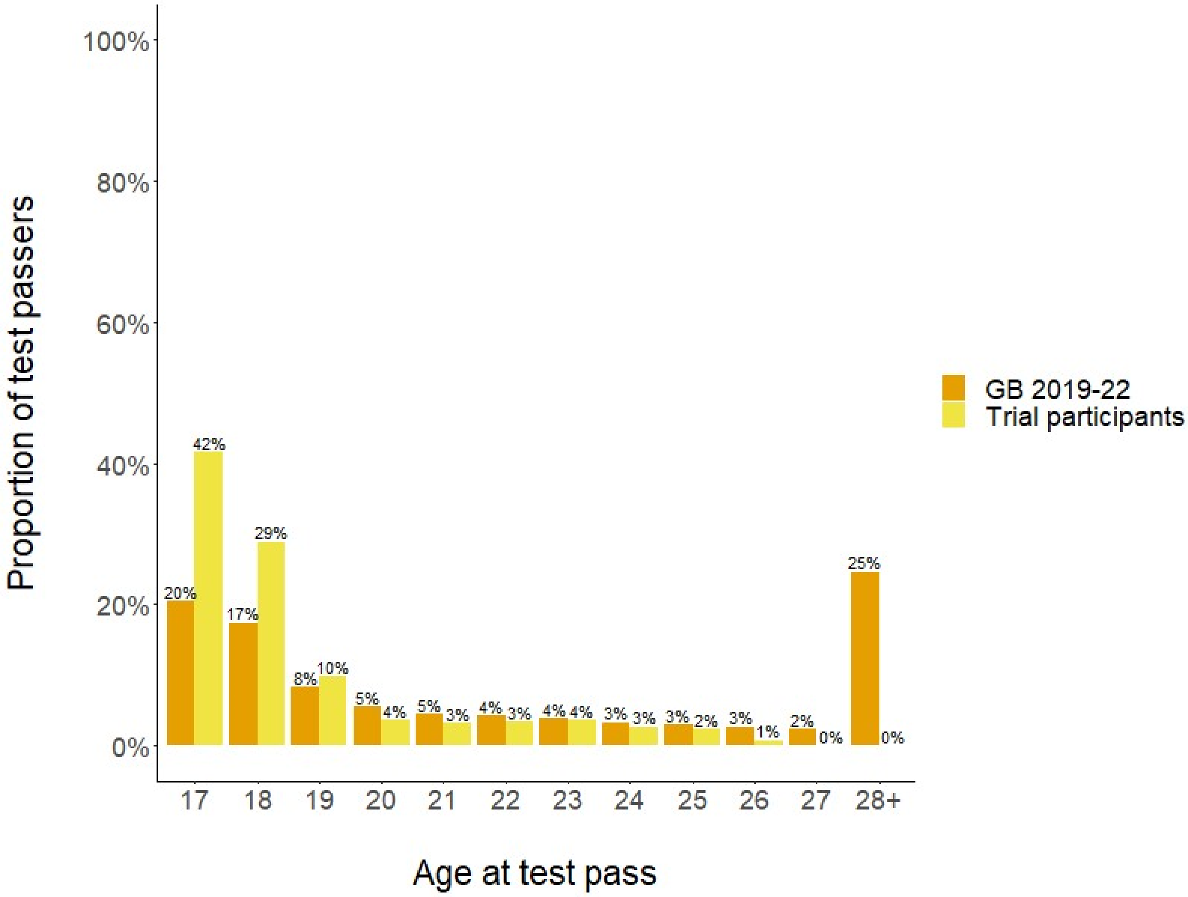


Figure 4-6: Proportion of test passers in Great Britain (2019-22) and trial participants by age at test pass (N=3,407)

There was a much higher proportion of 17 and 18-year-olds (at test pass) among the trial participants (71% combined) than seen across all test passers in Great Britain (38% combined after rounding). One quarter of all test passers in Great Britain were aged 28 or more.

These differences were expected given that the trial sample only included participants aged 16-24 at registration and then 17-27 at test pass. It means that findings should only be generalised to this age group.

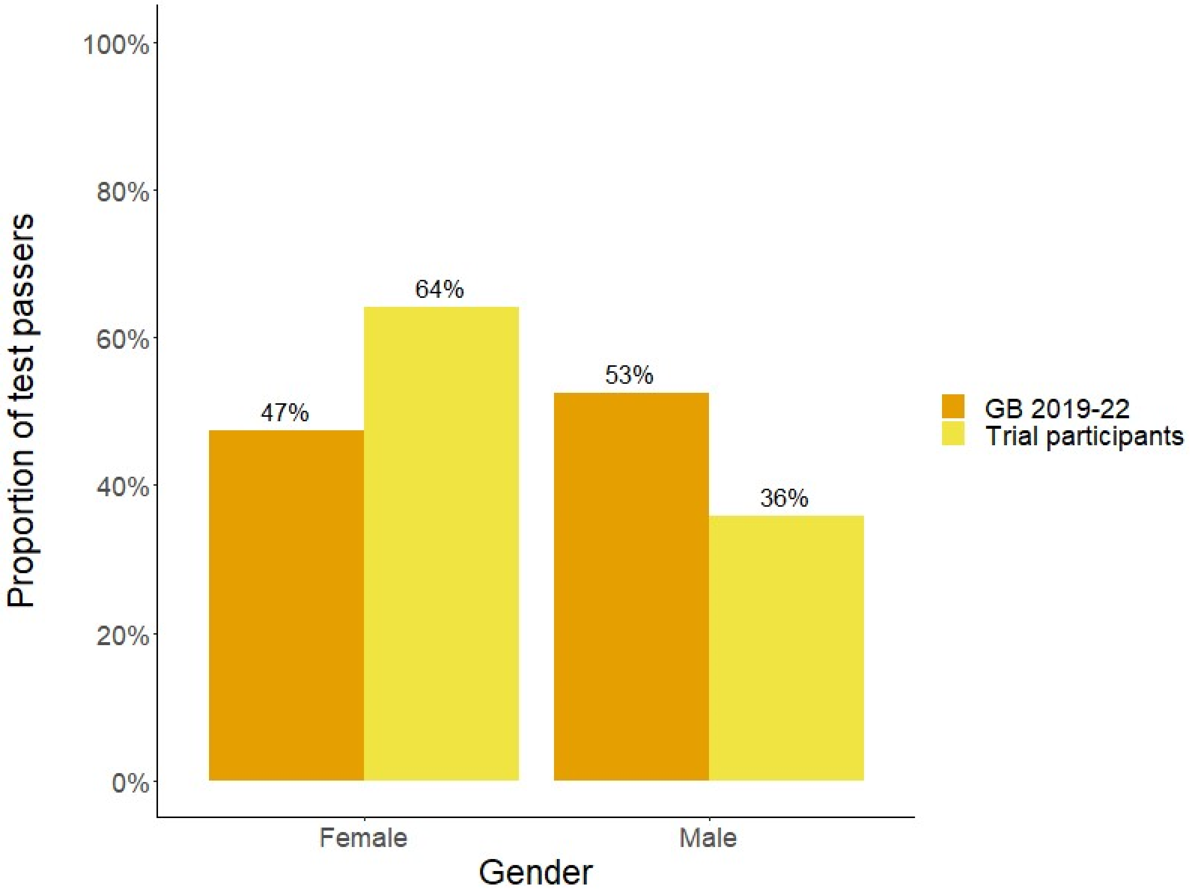


Figure 4-7: Proportion of test passers in Great Britain (2019-22) and trial participants by gender (N=3,407)

Figure 4-7 shows the difference in gender distributions between the trial sample and test passers nationally. The trial sample includes a higher proportion of females (64%) compared with test passers across Great Britain (47%). This is typical for studies of this type (for example Wells et al. (2008) had 63% females); the study was not reliant on achieving a representative gender balance, meaning that findings cannot be generalised to the population of young and novice drivers in Great Britain without some caution.

No comparison was done between the trial sample and test passers nationally for SDI or education status; this is because these data are not readily available for the national test passer population.

4.3.3 Spill-over and contamination checks

Checks were undertaken to ensure that participants in each group had not been affected by the interventions applied to any other groups (spill-over) or other known external factors (contamination) – in this case whether or not they had a telematics-based insurance policy.

The original intention was that checks for spillover would need to be carried out through a question in the 12 months post-test survey, therefore participants were asked whether they had heard of, or engaged with, any of a list of interventions (those in this trial). During the development of the interventions however, it became apparent that access to the interventions could be controlled much more closely than originally expected. From the control over access, and supported by the answers to the question, spill-over is extremely unlikely to have been an issue.

Participants were also asked at the 12-month time point whether they had a telematics insurance policy, in order to be able to control for any differences in this variable between the groups; previous work (Helman et al., 2017) had shown that having such a policy is associated with a greater number of self-reported collisions. Between 35% and 37% in each group had a telematics insurance policy. The proportions for the groups were not statistically significantly different (p = 0.48). This indicates that it is unlikely that any bias was introduced into the trial from contamination through telematics insurance policies.

4.3.4 Learning and driving conditions

This section explores the conditions in which trial participants learned to drive and the types of driving they did in the first year after passing their test. Despite some of these measures being targeted by the interventions (and so being findings relevant to the aims of some of the interventions), there were no notable differences identified between groups in either learning or post-test driving conditions.

4.3.4.1 Time spent learning

In the test pass survey participants were asked to estimate how many hours they had spent driving with an ADI and with another supervising driver (e.g., friends or family) before they passed their test. Table 4-4 presents the mean number of hours participants in each group reported that they spent learning with an ADI, with another supervising driver, and in total.

Trial participants reported spending an average of 71.5 hours driving before passing their test; 40.4 hours learning with an ADI and 31.0 hours practising with another supervising driver.

This was similar across all groups. The logbook intervention was designed to increase the number of hours of practice, but due to the low engagement no effect could be detected in this analysis, which included participants whether or not they engaged. The per-protocol analysis (see section 4.5.2.1) explores this in just those people who engaged with the logbook intervention.

Table 4-4: Mean (and standard deviation) of hours spent learning to drive with an instructor and with a supervising driver, by group (N = 3,407)

		Group				
		Control	Logbook	Hazard perception training	Education course	All participants
Mean (and standard deviation) hours spent learning to drive…	With ADI	40.3 (25.7)	41.1 (28.0)	40.7 (30.0)	39.5 (27.5)	40.4 (27.8)
	With other supervising driver	32.1 (38.3)	30.9 (38.8)	29.3 (36.9)	31.8 (40.2)	31.0 (38.5)
	In total	72.4 (45.9)	72.0 (47.7)	70.0 (47.3)	71.3 (49.0)	71.5 (47.4)

4.3.4.2 Learning conditions

In addition to answering questions about the number of hours spent learning, participants were asked to estimate the proportion of their learning mileage they spent driving in certain conditions including on different road types, with additional passengers (not including the supervising driver), in the dark, and in wet conditions. The proportion spent learning in each condition has been averaged across the participants in each group.

Figure 4-8 shows the proportion of learning miles participants in each group spent driving in different conditions. All three intervention groups and the control group show very similar results.

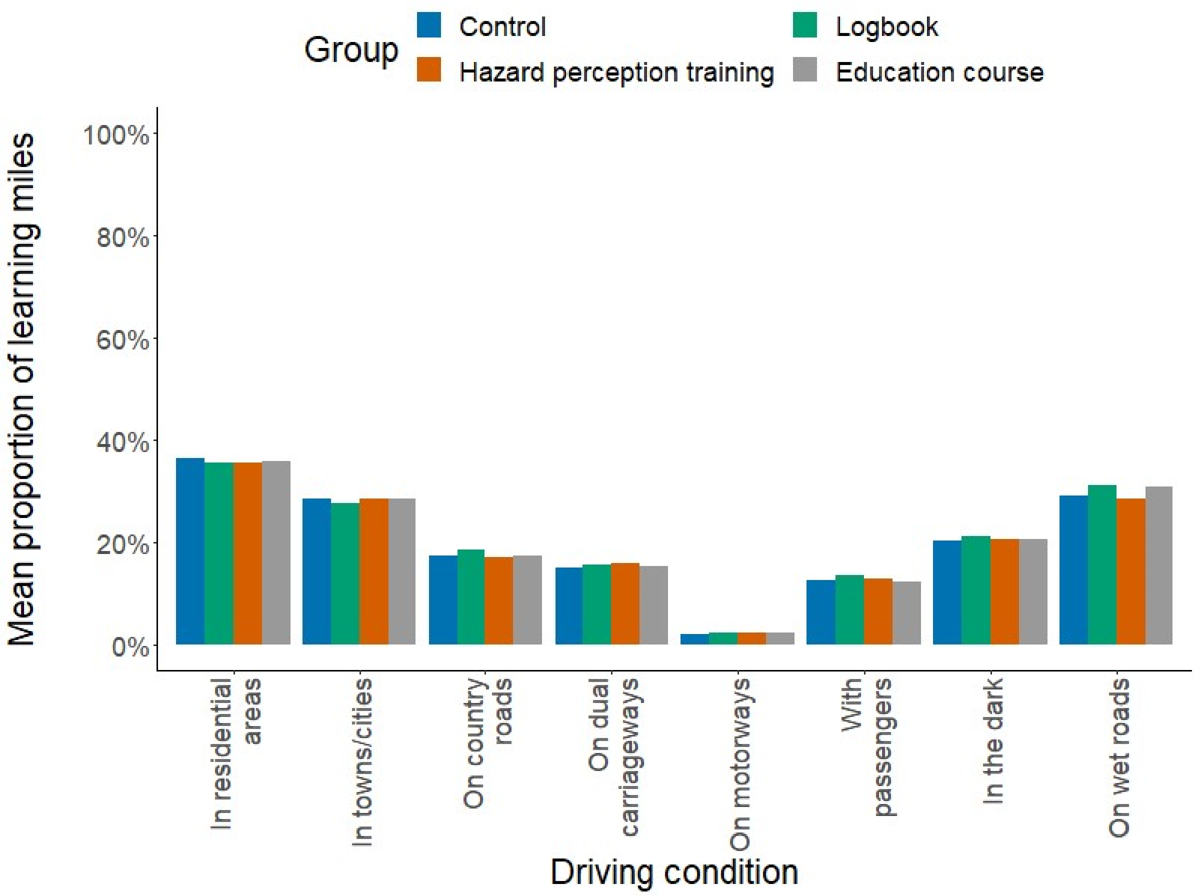


Figure 4-8: Proportion of learning miles spent driving in different conditions by group (N=3,407)

Of the different road types asked about, participants spent the most time learning in residential areas (36% of learning hours) and the least on motorways (2% of learning hours). This is to be expected, because residential areas are typically quieter and more suitable for learning and practising driving, especially at the beginning, and in particular for learning manoeuvres. Note also that motorway driving is not included in the practical test.

Participants spent 13% of their learning hours driving with passengers additional to their supervising driver. The proportions of learning hours spent driving in the dark and on wet roads were 21% and 30% respectively. The only statistically significant difference ($p<0.01$)

found was for proportion of time learning in the wet, with logbook and education groups being very slightly higher than the hazard perception training group, with a small effect size. There is no clear reason for this difference, given the very low engagement with the logbook intervention.

4.3.4.3 Mileage since test pass

Table 4-5 shows the mean number of miles driven in each reporting period and the total driven during the first twelve months since passing the driving test, by group. This analysis only includes the 3,292 participants answering all the surveys.

Participants drove an average of 5,621 miles in the first twelve months after passing their test. Those in the logbook group had the highest mean annual mileage (5,845 miles) and those in the education course group had the lowest (5,635 miles). Statistical testing showed no significant difference in the distributions of mileage between the four groups (p=0.55 – see Appendix C for a description of p-values and effect sizes).

Table 4-5: Mean (and standard deviation) of total miles driven by reporting period and in first twelve months since test pass by group (N=3,292)

		Group				
		Control	Logbook	Hazard perception training	Education course	All participants
Month 1-3	Mean (and standard deviation)	1,345 (1,533)	1,383 (1,604)	1,222 (1,426)	1,304 (1,458)	1,318 (1,514)
Month 4-6	Mean (and standard deviation)	1,424 (1,543)	1,530 (1,656)	1,551 (1,746)	1,387 (1,545)	1,476 (1,625)
Month 7-12	Mean (and standard deviation)	2,812 (2,396)	2,932 (2,557)	2,761 (2,421)	2,778 (2,339)	2,826 (2,436)
Miles driven in 12 months since test pass	Mean (and standard deviation)	5,582 (4,373)	5,845 (4,716)	5,534 (4,563)	5,470 (4,331)	5,621 (4,507)

4.3.4.4 Post-test driving conditions

At the end of each post-test reporting period (months 1-3 post-test, months 4-6 post-test, and months 7-12 post-test), participants were asked about the conditions in which they had driven during that reporting period. The questions were the same as those asked about learning conditions, but also included a question about the proportion of mileage spent driving for work (not including commuting). The analysis presented in this section only includes the 3,292 participants who completed all the surveys.

Figure 4-9 shows the proportion of miles driven in different conditions since test pass for participants in each group. The proportions reported at each survey point have been weighted based on reported mileage for each person and then averaged across the participants in each group.

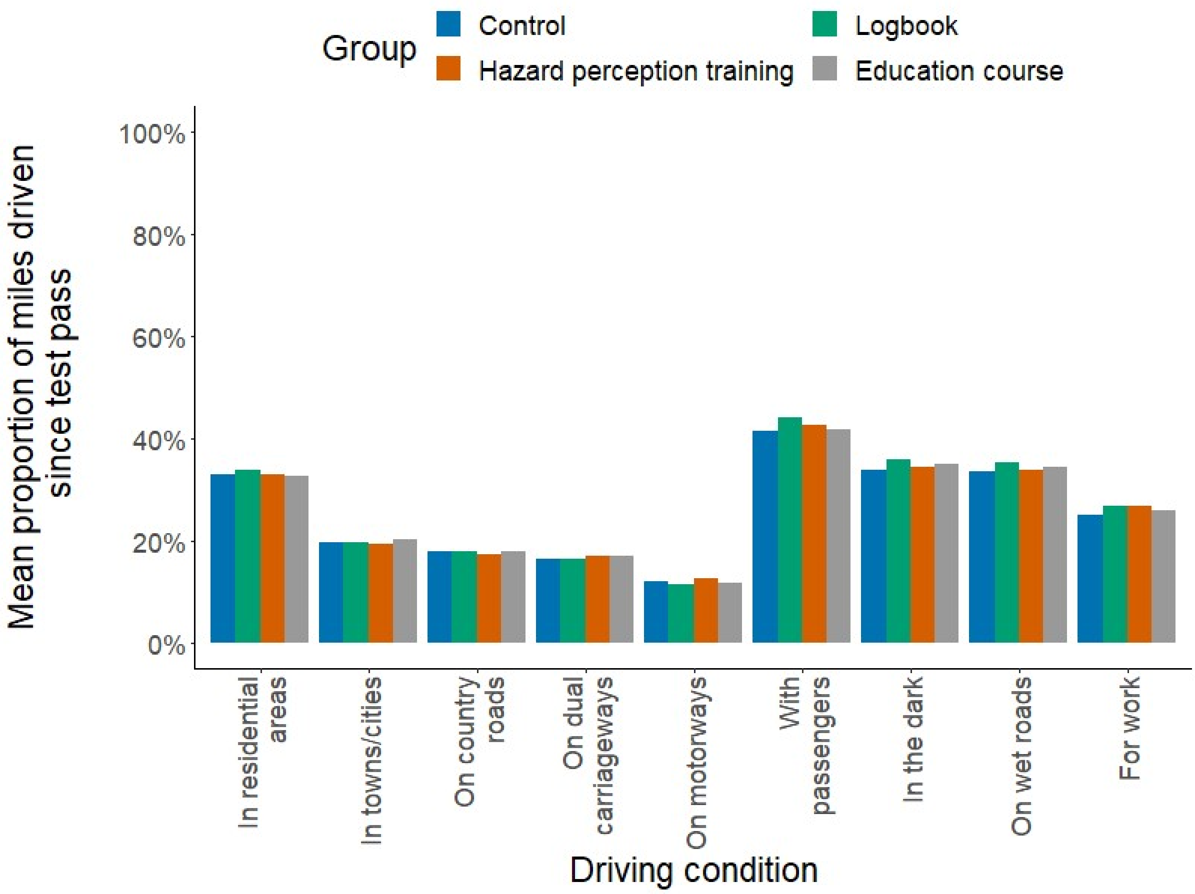


Figure 4-9: Proportion of total miles driven in 12 months since test pass in different conditions by group (N=3,292)

Figure 4-9 shows that, out of the different road types, participants most commonly drove in residential areas after passing their test, and motorways accounted for the smallest proportion of post-test mileage. This pattern is similar to that reported by participants whilst they were learning to drive (Figure 4-8) but the difference in proportions between motorways and other road types is smaller than for the learning phase.

The proportion of post-test mileage driven with passengers (43% across all groups) was much higher than seen pre-test (13%, see section 4.3.4.2). The proportion of driving done in the dark after participants passed their test (35% across all groups) was also higher than before they passed (21%) but the proportion of driving on wet roads was more similar before and after passing (30% and 34% respectively).

There were small differences between groups. For example, the control group had a slightly lower proportion of mileage with passengers, in the dark, on wet roads and driving for work. However, the only difference that was statistically significant was the proportion of driving with passengers between the control group and the logbook group (*p*=0.03). Given the extremely low engagement with the logbook intervention it is likely that this change is unrelated to the intervention itself.

Figure 4-10 and Figure 4-11 show how the proportions of post-test mileage in different road types and conditions changed across the three reporting periods. All four groups showed similar results.

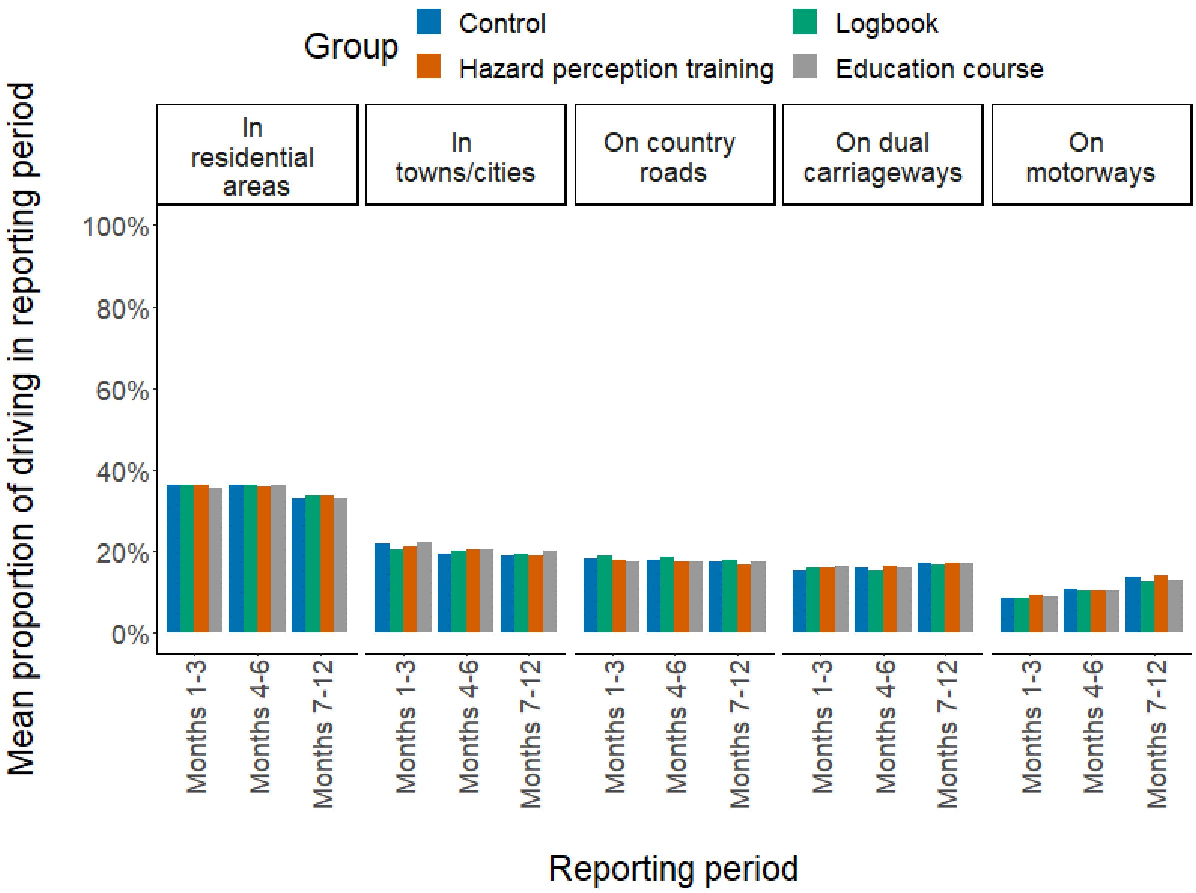


Figure 4-10: Proportion of miles driven in each reporting period on different road types by group (N=3,292)

Figure 4-10 shows that the proportion of post-test mileage driven by participants on different road type stayed broadly similar across the three reporting periods in their first year of driving, with the exception of the proportion of time spent driving on motorways. As time since test pass increased, participants reported an increase in the proportion of their driving that they did on motorways from 9% during the first three months post-test to 13% in the final six months of their first year of driving (proportions across all groups).

There were also small decreases over the three reporting periods in the proportions of mileage driven in residential areas, in towns and cities, and on country roads, although this does not necessarily mean that participants drove less in these locations during the reporting periods later in their post-test driving. Rather, it is likely that the increased proportion of motorway driving in later reporting periods meant that driving on residential and urban roads correspondingly accounted for a slightly smaller proportion of total mileage.

There were no noticeable differences between groups in the proportion of mileage driven on the different road types. Statistical tests showed no significant differences.

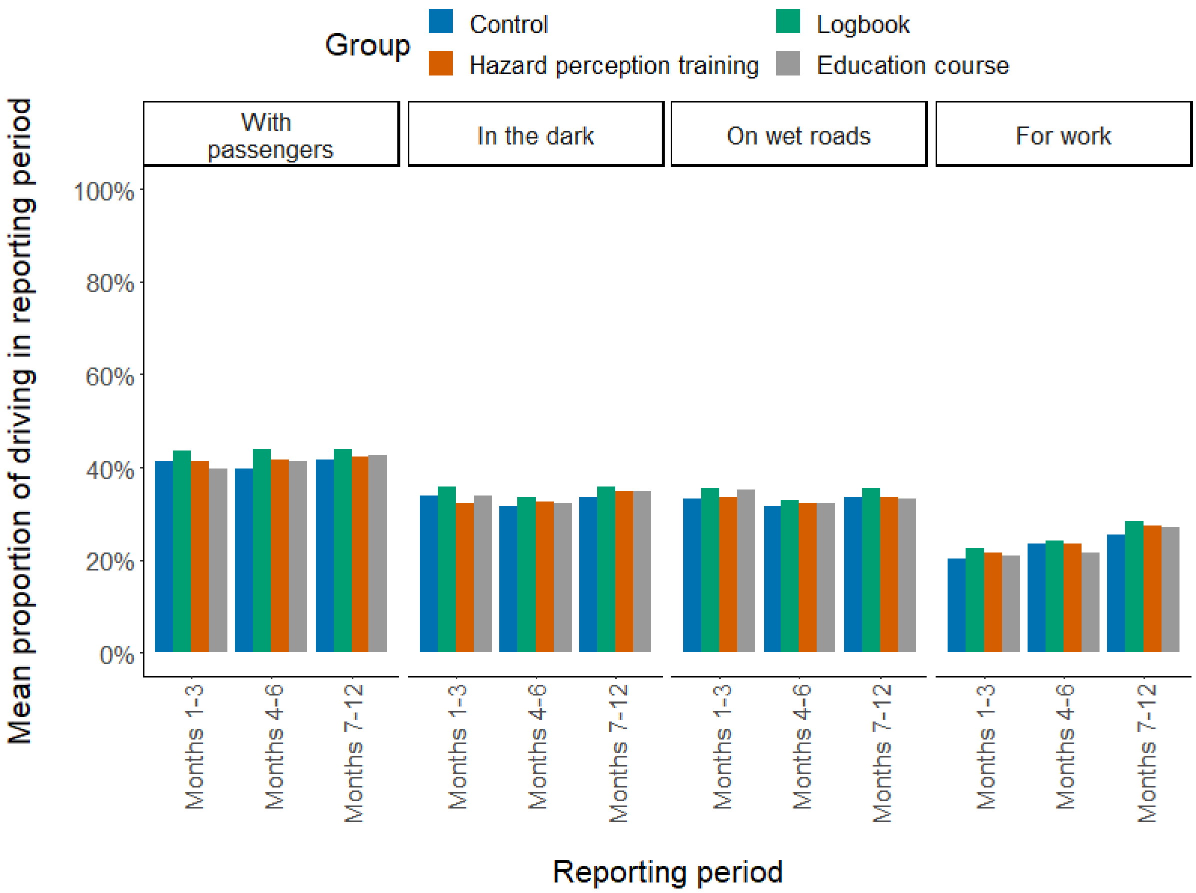


Figure 4-11: Proportion of miles driven in each reporting period with passengers, in the dark, on wet roads, and for work; split by group (N=3,292)

Figure 4-11 shows that the proportion of mileage driven with passengers, in the dark, and on wet roads was broadly similar across the three reporting periods, but the proportions of post-test mileage driven for work differed more across reporting periods.

Participants reported an increase in the proportion of their driving which was done for work (any driving for work purposes excluding commuting to and from their usual place of work) across the three reporting periods; 21% of mileage (across all participants in all groups) was driven as part of work in participants' first three months after passing their test and this increased to 27% in the final six months of their first post-test year.

The logbook group showed some significant differences compared to other groups: more passengers at 6 months compared with the control group (p<0.01); more driving in the dark at 3 months and compared with the hazard perception training group (p<0.01); and more driving on wet roads at 12 months compared with the education group (p=0.02) and hazard perception training group (p=0.02). These effects were very small and, given the very low engagement with the logbook intervention (around 3% of participants engaged), they seem unlikely to be due to the intervention. It would be speculative to suggest why this effect did occur, although the per-protocol analysis in section 4.5.3.1 looks at the same variable just for engagers.

4.3.5 Impact of COVID-19 pandemic

This section explores the impact of the COVID-19 pandemic on participants' learning to drive process and post-test mileage; exploratory analysis showed that the pandemic had an impact on both.

4.3.5.1 Impact of COVID-19 pandemic on test pass date

Of the 3,407 participants in the trial sample, 1,403 (41%) passed their test before 20th March 2020 (the start of the first lockdown) and 2,004 (59%) passed their test after this date. Henceforth, the period prior to 20th March 2020 is labelled as "Pre-COVID-19" and the period after this date is labelled as "Post-COVID-19".

Figure 4-12 shows the distribution of participants by date of passing their practical driving test.

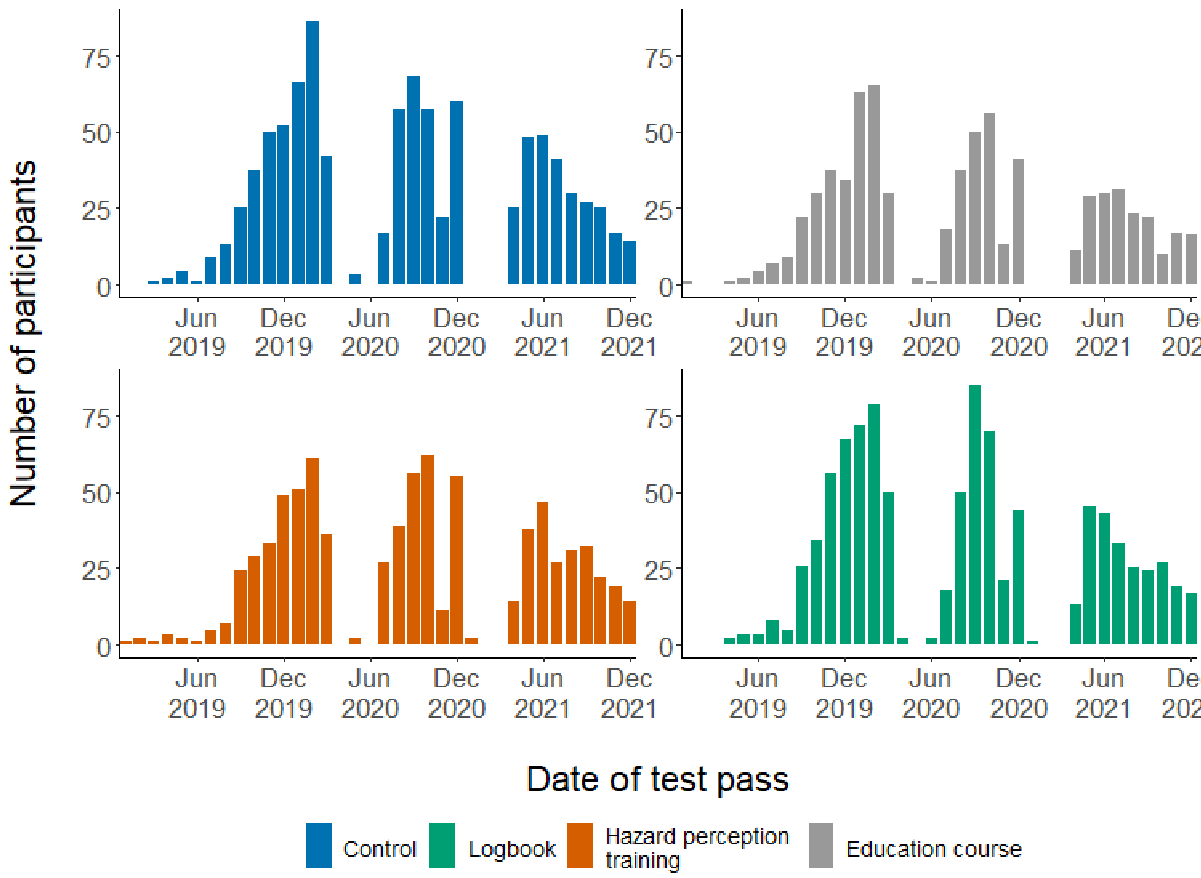


Figure 4-12: Distribution of participants by date of test pass (N=3,407)

There was no significant difference in the distribution of date of test pass between groups. The proportion of participants in each group that passed their test prior to the start of the pandemic ranged from 35% to 40%.

4.3.5.2 Impact of COVID-19 pandemic on learning to drive

In the test pass survey, participants were asked if they had experienced any impacts on their learning to drive process due to the COVID-19 pandemic. These included impacts such as postponement or cancellation of their theory or practical tests, being unable to book their tests when they were ready to, having more or less driving experience with their instructors or parents and having the same amount of experience but taking a longer time due to the pandemic.

The results are shown in Figure 4-13. Participants were allowed to select all options that applied to them and therefore, the proportions do not add up to 100%.

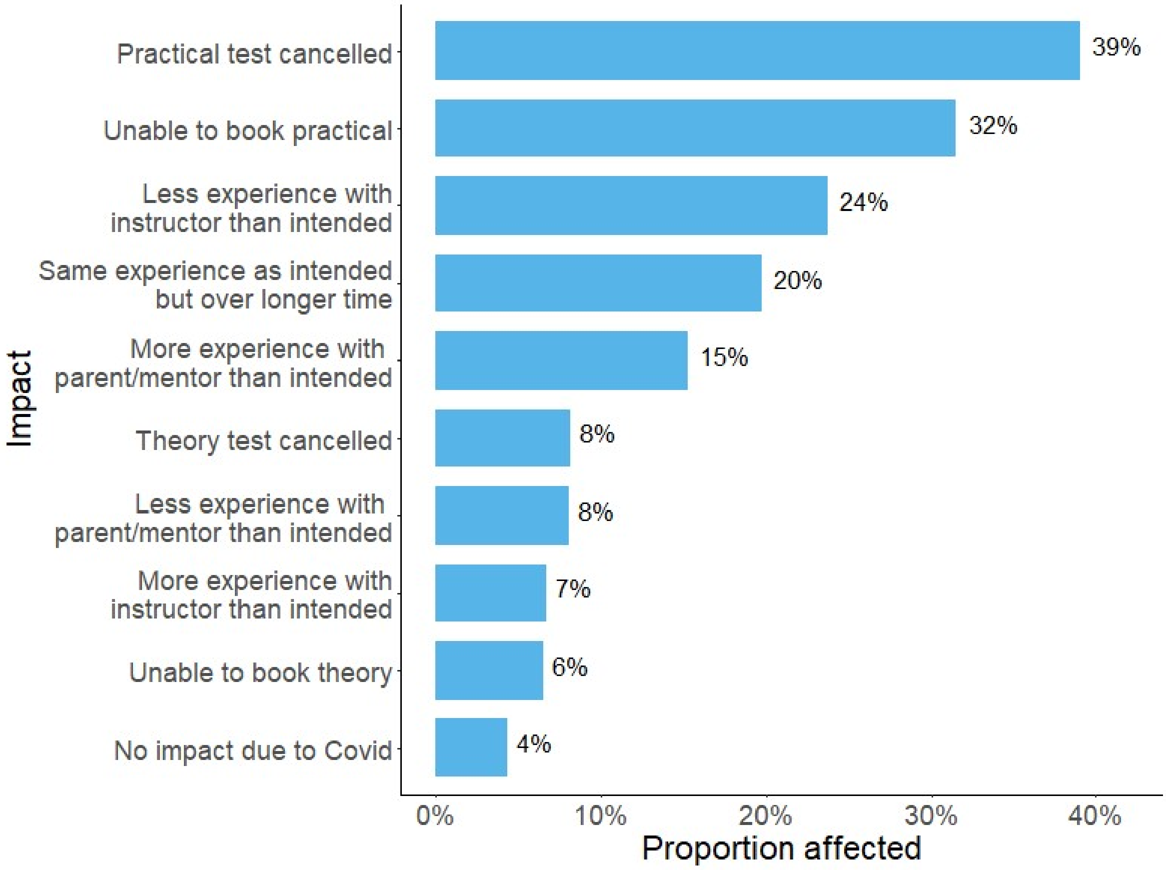


Figure 4-13: Impact on the learning to drive process due to the pandemic (N=3,407)

The top three impacts reported by participants were their practical driving test being postponed or cancelled (roughly 39% of respondents in each group), not being able to book their practical driving test when they were ready to (roughly 32% of respondents in each group) and having less experience with their driving instructor (roughly 24% of respondents in each group). The results did not vary by group.

The above responses were converted into a numerical score for easier comparison during the intention to treat (ITT) modelling (see section 4.4). A score of 0 was given if participants said that the pandemic had no impact on their learning to drive process. The impact factors are shown in Table 4-6. Factors such as being unable to book their theory or practical tests, or tests being postponed or cancelled increased the participant's score by 1 as it impacted their learning to drive process in a negative manner. However, if participants reported having more driving experience with their instructor or parent then this reduced their score by 1, on the grounds that this impact was positive in nature.

Table 4-6: Calculating a score for the impact of COVID-19 pandemic on learning to drive

Factor	COVID-19 pandemic impact on learning to drive score modifier
No impact on learning to drive process due to the pandemic	0
Practical test booked which was postponed or cancelled	1
Theory test booked which was postponed or cancelled	1
Unable to book practical test when ready to	1
Unable to book theory test when ready to	1
Less experience with instructor than intended	1
Less experience with parent than intended	1
More experience with instructor than intended	-1
More experience with parent than intended	-1
Took longer to get the amount of experience intended	1

This resulted in a COVID-19 Impact Score for each participant, where a higher value denotes a greater negative impact of the pandemic on the learning to drive process and a lower value denotes a lower negative impact of the pandemic on the learning to drive process. A negative value suggests that the pandemic actually had a positive impact on the participants' learning to drive process (for example by helping them get more practice).

As this question did not apply to those who passed their test prior to the pandemic, all participants in this group were given a score of 0. Participants who passed their test after the start of the pandemic had a score ranging between -2 and 7, with the average score for the group being around 2.0 (standard deviation of 1.3), based on the scoring system for impact factors noted above.

4.3.5.3 Impact of COVID-19 pandemic on mileage after passing their test

This section explores the impact of the pandemic on participants' mileage after they passed their practical driving test. Figure 4-14 and Table 4-7 show the summary of annual mileage reported by participants pre and post the start of the pandemic.

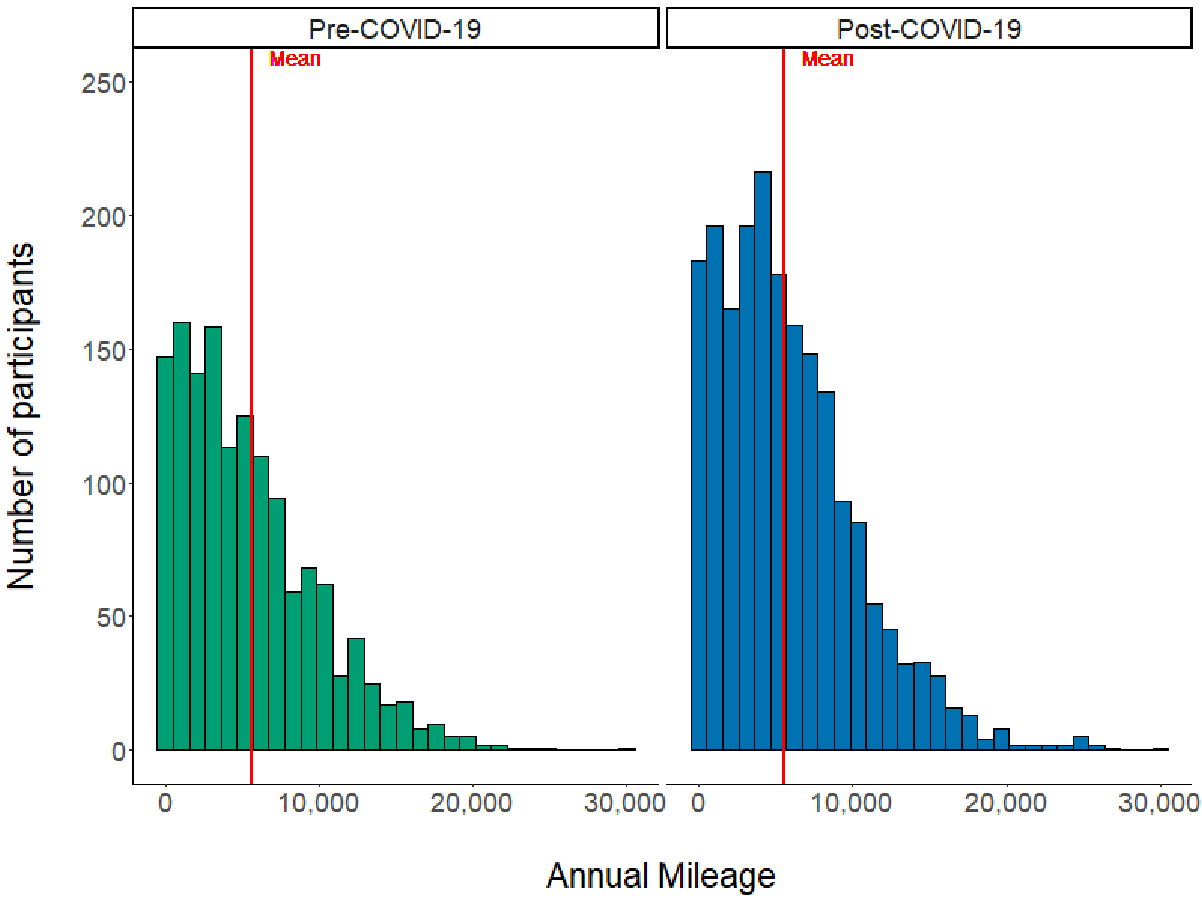


Figure 4-14: Distribution of annual mileage by test pass date (pre or post pandemic) (Pre-pandemic N=1,403, post-pandemic N=2,004)

Table 4-7: Summary of annual mileage by test pass date (pre or post pandemic) (N=3,407)

		Pre-COVID-19	Post-COVID-19	All participants
Miles driven in 12 months since test pass	Mean (Standard deviation)	5,399 (4,464)	5,755 (4,527)	5,609 (4,504)
Total number of participants		1,403	2,004	3,407

Participants who passed their driving test prior to the start of the pandemic reported lower annual mileage (5,399 miles) than those who passed their test after the pandemic began (5,755 miles). This difference was statistically significant (*p*=0.02 with a small effect size of 0.05. This result is expected as those who passed their test prior to the start of the pandemic would have been impacted by the various lockdown periods after test pass to a greater extent.

4.3.5.4 *Reasons for not driving*

Some participants did not drive in each reporting period. This is summarised in Table 4-8. Participants who did not drive at all in the first twelve months (122 participants) were excluded from the statistical modelling (conducted in section 4.4).

Table 4-8: Number of participants who did not drive in each reporting period

Reporting period	Number of participants who did not drive	Proportion of participants who did not drive
Months 1-3	314	10%
Months 4-6	291	9%
Months 7-12	200	6%
First 12 months	122	4%

At each reporting period, participants were asked for specific reasons that may have limited their driving in the reporting period. This included reasons such as not having access to a vehicle, not needing to drive due to change in where participants lived or worked, not needing to drive due to health conditions or to save money, being banned from driving or having lost their licence and due to the COVID-19 pandemic. Figure 4-15 presents the trends of the top four reasons for limitations on their driving over time, for participants who did not drive in each reporting period. As participants would select multiple reasons, the proportions do not add to 100%.

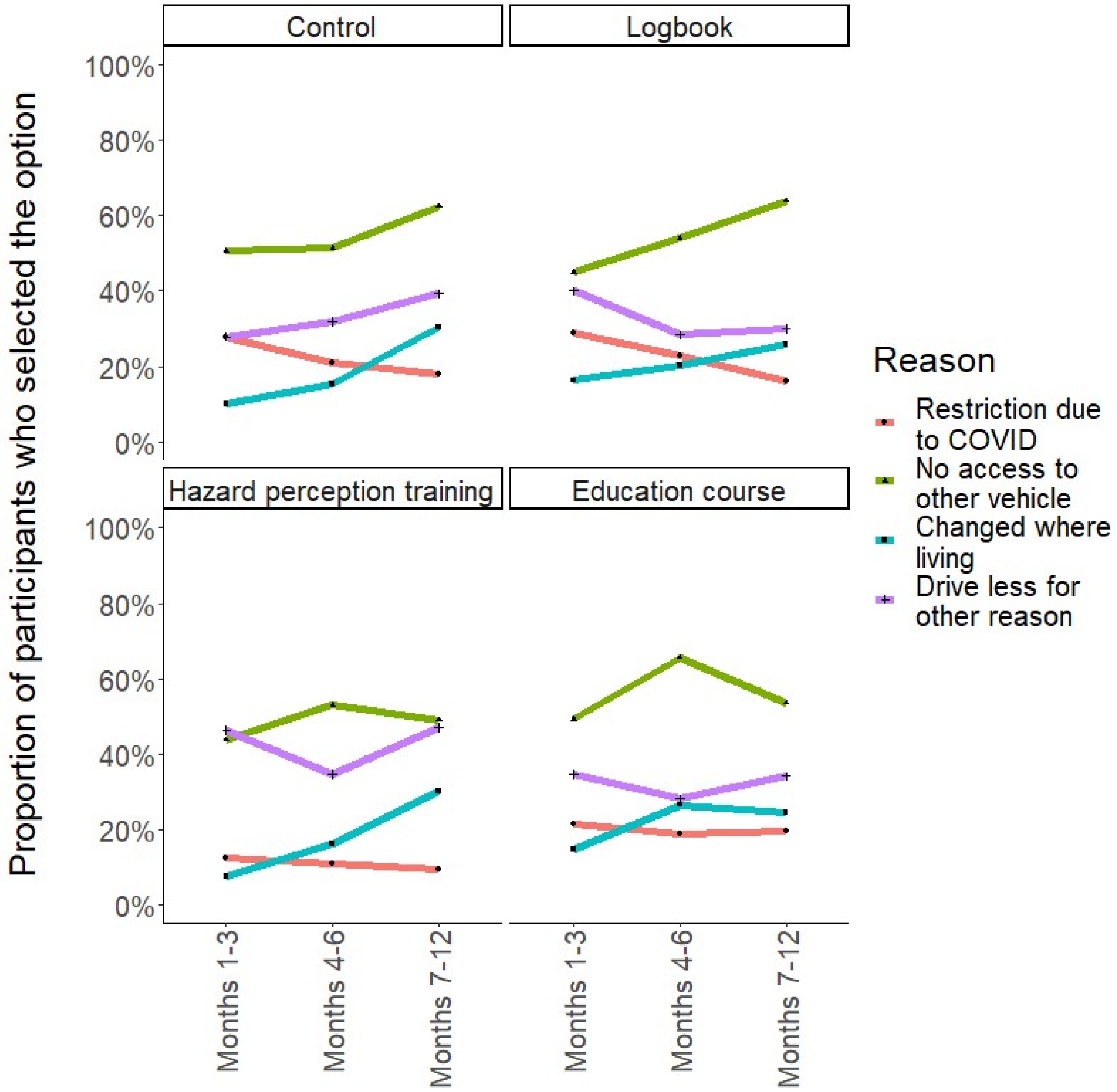


Figure 4-15: Top four reasons for not driving by reporting period

Not having access to another vehicle was the most common reason for not driving in almost all reporting periods shown by the green line in the figure. The effect of the pandemic on driving seems to decrease over the twelve months since test pass; the only significant difference between the control and intervention groups for this variable was seen in months 1-3 for the control and hazard perception training groups (p=0.03). This difference had a small effect size.

4.3.6 Exploratory analysis of collisions

At each reporting period (months 1-3, months 4-6 and months 7-12) participants were asked how many collisions they had been involved in over the duration of that period. The responses from each reporting period were combined to obtain the total number of collisions involved in at 12-months since test pass. The descriptive analysis shown in this section excludes participants from the trial sample who did not drive at all in 12 months and aligns with data used to develop the statistical model in section 4.4.

4.3.6.1 Exploring collisions by age, gender and mileage

Previous studies (for example, Maycock et al., 1991; Forsyth et al., 1995; Wells et al. 2008) examining the collision rate of young and novice drivers found that the number of reported collisions was related to age, gender, and exposure. More recent studies (Helman et al., 2017) found no clear association between collisions and gender, suggesting that the gender differences have reduced over time when using this kind of measure (self-reported collisions, which tend to be dominated by 'damage only' collisions). Figure 4-16 shows the average number of self-reported collisions per person by age and gender for the 12 months post-test driving. The number of collisions per person does seem to spike at 20 and 21 years of age. The difference by gender in the collisions per person for 25 and 26 year olds can be explained by the small sample sizes for those ages.

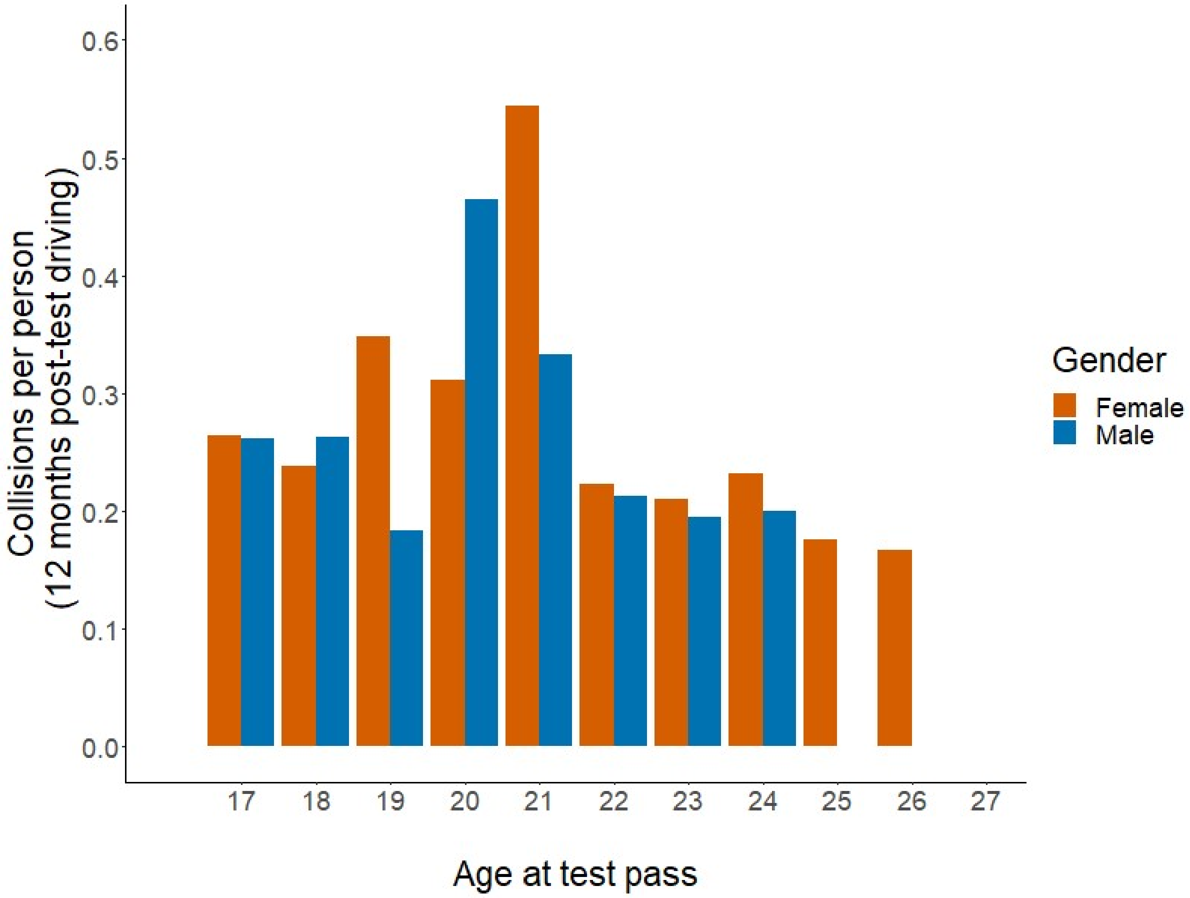


Figure 4-16: Average number of self-reported collisions per person by age and gender (excluding those who did not drive in 12 months, N=3,278)

The lack of an obvious difference by gender (average collisions per person for males = 0.25 and females = 0.27) appears to support the 'closing' of gender differences over time in studies in Great Britain. The apparent differences by age (average collisions per person for 17-year-olds = 0.26, 20-year-olds = 0.37, and 24-year-olds = 0.22 – note the apparent increase for 21-year-old females and 20-year-old males also) is examined in the statistical modelling, although note that the number of participants is much smaller than expected and therefore this measure will be potentially more prone to random fluctuations.

Figure 4-17 shows the distribution of mileage by the number of collisions reported by participants. This has been represented by a violin plot (shown by the outer shaded region) and boxplot. The violin plot shows the density of the data at different values and the boxplot shows the inter-quartile ranges (25^{th} and 75^{th} percentile) and the median value (50^{th} percentile or 'median average'). For instance, the violin plot shows that participants reporting no collisions generally reported a lower annual mileage with the most around 2,500; for participants reporting three or more collisions, the distribution of mileage reported was more spread.

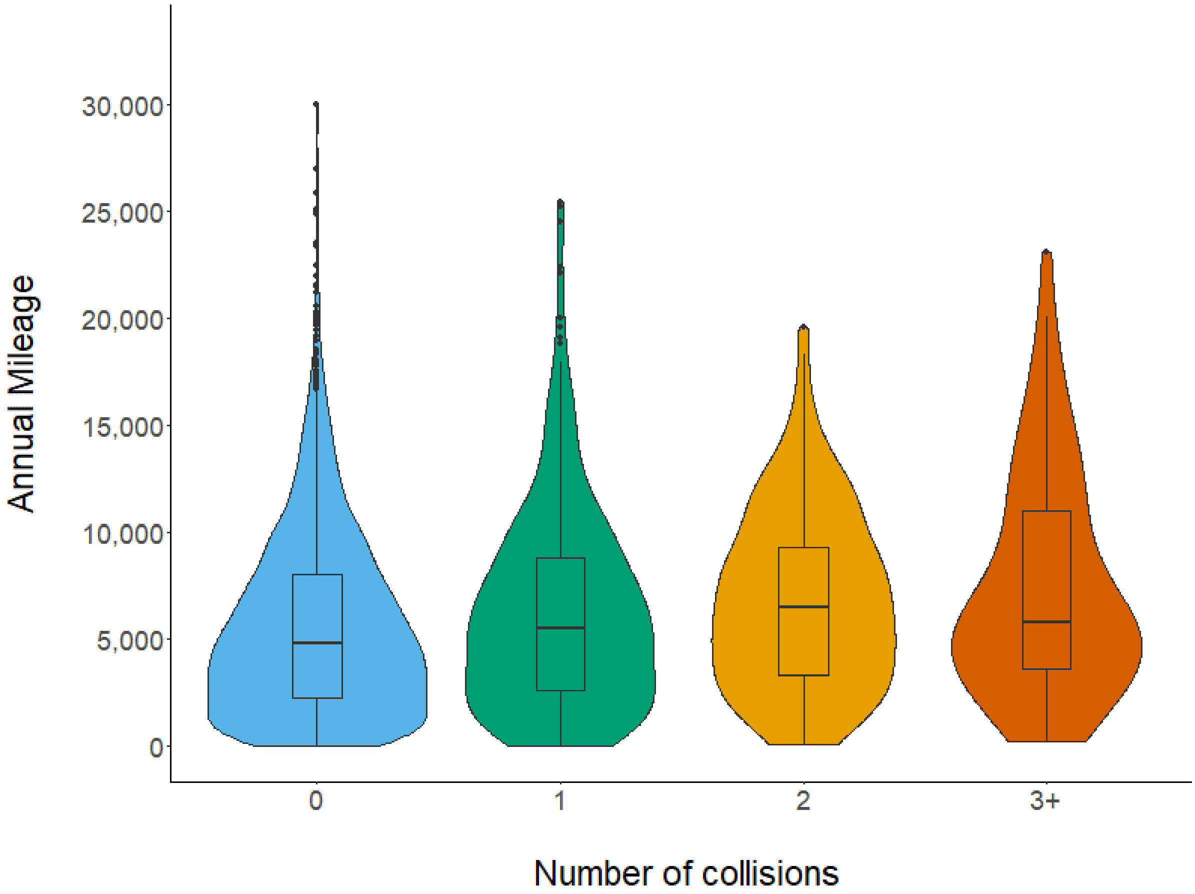


Figure 4-17: Distribution of mileage by the number of collisions (excluding those who did not drive in 12 months, N=3,278).

Participants who had higher annual mileage reported a higher number of collisions on average. The average annual mileage for those who reported having three or more collisions (7,429 miles) was significantly different (p<0.01) compared with those who reported having no collisions (5,672 miles). As a result, mileage was considered an important factor to account for in the collision modelling in section 4.4.

4.3.6.2 Exploring collisions by intervention group

Figure 4-18 shows the proportion of collisions reported by group. The majority (around 82%) of the participants across all groups reported that they had not been involved in any collisions. Roughly 12% in each group reported being involved in one collision. There was no significant difference (*p*=0.58) in the number of people reporting at least one collision between groups.

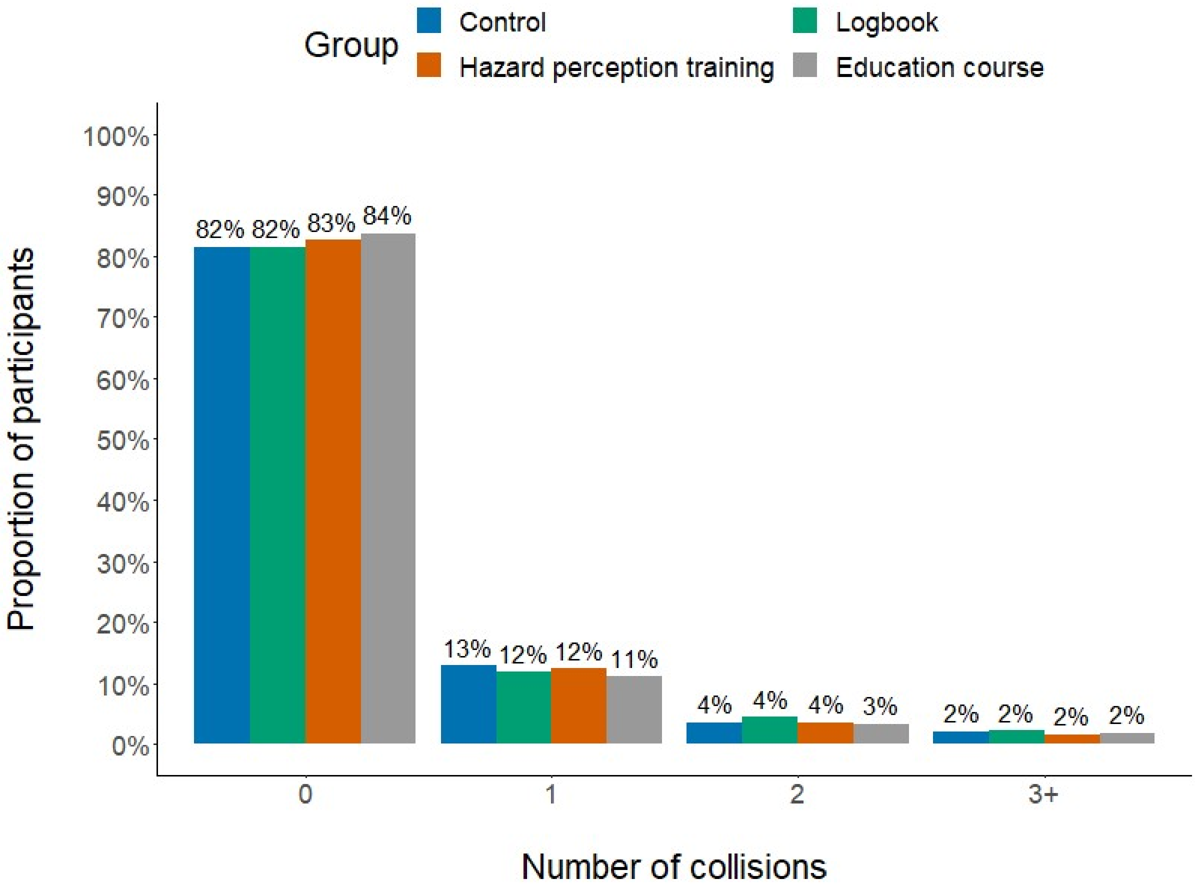


Figure 4-18: Number of self-reported collisions by group (excluding those who did not drive in 12 months, N=3,278)

Table 4-9 shows collision rate per thousand miles driven by each group. The results show there was no apparent difference between the groups. No statistical testing was undertaken on these rates. Statistical modelling of the effect of the interventions on collisions, accounting for various potential confounding variables in addition to mileage, is presented in section 4.4.

Table 4-9: Collision rate per thousand miles of driving by group

	Control	Logbook	Hazard perception training	Education course	All participants
Total number of participants	914	910	769	685	3,278
Total number of collisions reported in 12 months	240	248	185	161	834
Total miles driven in 12 months	5,284,898	5,512,581	4,458,340	3,852,453	19,108,272
Average mileage driven per person	5,782	6,058	5,798	5,624	5,829
Collision rate per thousand miles	**0.045**	**0.045**	**0.041**	**0.042**	**0.044**

4.3.7 Attrition analysis

Section 3.2.2 described how participants were randomly assigned to groups (with a slight weighting towards the logbook intervention group); this ensured that the characteristics of the sample across the intervention groups were similar.

As participants dropped out of the study, the samples for the groups changed. This means that the characteristics of the groups could change in comparison with the characteristics of the registration sample. An attrition analysis was therefore undertaken to investigate how the groups changed over time as participants dropped out, and whether the way the characteristics of the samples changed were consistent over time across the groups. In other words, the analysis checked for any bias introduced by different types of people dropping out of different groups.

Overall, there were changes in the characteristics of the sample over time as participants dropped out. However, these changes were consistent across the groups. There were no noticeable differences in the behaviour of the characteristics for the groups. Therefore, there was no need to correct samples for differences between them.

4.4 Intention to Treat (ITT) analysis

The intention to treat (ITT) analysis analysed the effect of the interventions on collisions (relative to the control group) at the population level. It included participants who did, and did not engage with the interventions.

This section first examines the relationship between self-reported collision numbers and key driver variables, namely age, gender, mileage, and the impact of the COVID-19 pandemic in greater detail. Participants who did not drive at all in the first twelve months (129 participants) were excluded from the statistical modelling.

This model, independent of any intervention effect, forms the base model (shown in section 4.4.4) and can be used to hold certain factors constant when undertaking other analyses. Next, the intervention groups (control, logbook, hazard perception training, education course) are added to the base model forming the ITT analysis (results in section 4.4.5). This tests whether the group to which a participant was assigned (control, logbook, hazard perception training, education course) impacted on their collision risk at twelve months post-test, independently of their level of engagement with the intervention on offer. In addition to the models above, the effect of driver experience (effect of months since test pass) on the number of reported collisions was modelled.

The key insights from this analysis are as follows:

- Higher mileage, passing the test before the COVID-19 pandemic, and a higher age at test pass were all associated with greater collision risk.
- Mileage had a weaker association with collision risk for those who passed their test before the COVID-19 pandemic than for those who passed after.
- None of the interventions were associated with any change in collision risk.

4.4.1 Background

The hypotheses for the three interventions were as follows:

Logbook: A mobile-phone-based app that encourages greater and more varied on-road practice when learning to drive, provided to learner drivers aged 17-24 in Great Britain (for them to use if they wish) will result in a lower proportion of such drivers who have a crash in their first 12 months of post-test driving, relative to a control group of drivers who do not receive the opportunity to use the mobile-phone-based logbook app when learning.

Hazard perception training: A web-based hazard perception eLearning training package (three modules) provided to learner drivers aged 17-24 in Great Britain (for them to use if they wish) will result in a lower proportion of such drivers who have a crash in their first 12 months of post-test driving, relative to a control group of drivers who do not receive the opportunity to use the web-based hazard perception eLearning training package.

Education course: A classroom-based/online-based education course (five modules) provided to learner drivers aged 17-24 in Great Britain (for them to use if they wish) will result in a lower proportion of such drivers who have a crash in their first 12 months of post-test driving, relative to a control group of drivers who do not receive the opportunity to use the classroom-based/online-based education course.

Based on previous studies of this type (Maycock, Lockwood & Lester, 1991; Forsyth et al., 1995; Wells et al., 2008; Helman et al. 2017) one would expect to see some or all of the following have independent impacts on collision risk in the first year of post-test driving:

- Exposure: drivers who drive more would be expected to have higher collision risk
- Age: younger drivers would be expected to have higher collision risk
- Gender: in some of the previous studies males have had a higher collision risk than females (although this difference was reducing throughout the 1990s and 2000s, and was not present in the most recent analysis reported in Helman et al., 2017)

- Experience: drivers would be expected to have a higher collision risk in their early months of driving post-test than in their later months of driving.

There may also be other variables that need to be accounted for in the current dataset – specifically there are potential impacts of the COVID-19 pandemic to consider; these were explored.

4.4.2 Modelling method

The modelling method applied to the data in this study was a multivariate regression technique known as Generalised Linear Modelling (GLM). Multivariate regression techniques are designed to explore the relationship between the response variable (in this case the number of self-reported collisions at 12 months since test pass) and a number of exploratory variables (such as age and gender) on which the response variable is assumed to depend. In this study, the response variable, number of collisions at 12 months since test pass, is assumed to follow a negative binomial distribution. The decision of using a negative binomial distribution over a Poisson distribution was determined through likelihood ratio and overdispersion tests. This distribution is frequently used to model count data such as collisions and has been used in previous studies of similar nature (Wells et al., 2008; Helman et al., 2017).

The base model was developed by individually adding key driver variables to establish whether the addition has a statistically significant effect on the response variable (number of collisions). The key driver variables were identified from previous studies (section 4.4.1) on young and novice drivers. In addition to these variables, certain impacts of the COVID-19 pandemic were included in the model to explore the degree to which it may have impacted the number of collisions.

4.4.3 Impact of experience

Previous studies have shown that experience has a significant impact on the likelihood of a collision. Wells et al. (2008) modelled experience as the number of years the participant has been driving and the same process was followed in this study (using months rather than years). In order to model this, a repeated measures design was applied, and experience was coded as 0.25 for those who reported driving in the first six months and 0.75 for those who reported driving in the second six months since test pass.

Generalised Linear Models (GLM) assume that the data modelled are independently distributed (the response from one participant is independent of the other). However, in order to develop a model which includes experience over time as an explanatory variable, an approach which considers that some of the data are related is needed (as each participant responds to the surveys at multiple time points). Therefore, a mixed effects model was used to examine the relationship between driver experience and number of collisions. Mixed effects models are applied in settings where repeated measurements are made on the same participant, and they can also account for any missing values. It was assumed that the number of collisions followed a negative binomial distribution (similar to the base model using GLM).

The variables included in the model were mileage, age at test pass and gender along with the additional variable, experience. The model showed that experience was statistically significant ($p<0.01$) with a negative coefficient of 0.62 suggesting that the likelihood of a collision reduces as experience since test pass increases. This is what would be expected based on previous findings from studies of this kind, although the coefficient in this sample was higher than the coefficient of 0.40 seen in the novice arm of the study (Weekley et al., 2024a), something returned to in the discussion.

4.4.4 Base model

The base model for the reported number of collisions in the first 12 months after passing their practical driving test was of the form:

$\log(\text{Collisions}) = \beta_0 + \beta_1 \log(\text{Mileage}) + \beta_2 \text{Gender} + \beta_3 \text{Age} + \beta_4 \text{TestPassDate} + \beta_5 (\log(\text{Mileage})\text{:TestPassDate})$

where β_n are the coefficients to be estimated. The variables were included one by one in the following form:

- Mileage: This is number of miles driven by participants in the 12 months since test pass. This variable was included as a natural logarithm to account for the fact that the likelihood of a collision increases as mileage increases, but not in a simple linear manner.
- Gender: This was included as a categorical variable, where 0 was coded as female and 1 was coded as male.
- Age: This is age of the participant at the time they passed their practical driving test. A variety of functional forms were tested (logarithmic, exponential, inverse and categorical). A categorical approach with two groups, 17-21 and 22-27 years old, was found to be most appropriate fit for the data, where 0 was coded as the 17-21 group and 1 was coded as the 22-27.
- TestPassDate (pre- or post-pandemic): This was determined based on the date of test pass (before or after the start of lockdown in March 2020) as a surrogate for the impact of the COVID-19 pandemic on passing the test. Post-COVID-19 pandemic was coded as 0 and Pre-COVID-19 pandemic was coded as 1.
- Log(mileage):TestPassDate: This variable is the interaction between the mileage and test pass (pre- or post-pandemic) indicator.

No term for experience was included as the main analysis considered the first 12 months of driving as a whole; GLMs are not able to handle temporal variables well, and a decision was made to exclude experience from the base model on this basis. The coefficients for the base model are shown in Table 4-10.

Table 4-10: Coefficients for the base model (* denotes statistically significant coefficients)

Variable	Coefficient	Standard Error	*p*-value
Intercept	-3.916	0.550	<0.01(*)
Log(Mileage)	0.298	0.064	<0.01(*)
Gender: Male	-0.088	0.092	0.34
Age group: 22-27 years old	-0.357	0.143	0.01(*)
Test pass: Pre-COVID-19 pandemic	1.629	0.745	0.03(*)
Interaction between mileage and test pass	-0.164	0.088	0.06

The base model showed that:

- The log of mileage coefficient was statistically significant ($p<0.01$) with a positive coefficient of 0.30. This suggests that as mileage increases, the likelihood of a self-reported collision increases. Assuming a 17 to 21-year-old male who passed his test during the pandemic and therefore not activating the interaction term between mileage and COVID impact, the model suggests that if mileage is increased from 2,500 miles to 5,000 miles, then the likelihood of a collision is 1.23 times as high (an increase of 23% from 0.188 to 0.231). The effect of the interaction term being active on mileage is described when explaining the interaction term below.
- Unlike in earlier studies (Forsyth et al., 1995; Wells et al., 2008) but in line with the findings from the more recent DVSA study (Helman et al., 2017), gender was not shown to have a significant impact on the number of self-reported collisions. However, despite the non-significance of this variable, it was decided that the base model should include this variable in order to control for any effects of gender when considering the impact of the interventions in the ITT analysis (see section 4.4.5).
- The age at test pass group variable was shown to be a statistically significant predictor ($p=0.01$) of collisions. For a male participant who travelled 2,500 miles and passed their test during the pandemic, the likelihood of collision decreases by 30% from 0.188 to 0.132 when moving from the 17-21-year-old category to the 22-27-year-old-category. Note that in the novice arm, no effect of age was found on collisions. It is possible that because the participants in the learner arm were more variable in age at test pass than those in the novice arm (17-27 versus 17-24 respectively) the effect was detected as statistically significant in the former but not in the latter.
- The date of test pass (Pre-COVID-19 or Post-COVID-19) is statistically significant with a positive coefficient of 1.63 for those who passed their test before the pandemic ($p<0.01$). This suggests that those who passed their test before the pandemic began were up to 5.1 times more likely to be involved in a collision (a 410% increase) than those who passed their test after the pandemic began. This does not include the effect of the mileage and test pass interaction term (the effect of passing before the pandemic varied depending on mileage).

- The interaction between mileage and date of test pass (Pre-COVID-19 or Post-COVID-19) was close to being statistically significant but did not pass the significance threshold (p=0.06). This means that the coefficient of log(mileage) decreases by 45% if the participant passed their test before the pandemic. In a situation where a 17-21-year-old participant passed their test pre-COVID-19, the likelihood of a collision increases by 1.098 times (an increase of 10% compared with a 23% increase for the overall effect of mileage) when going from 2500 miles to 5000 miles driven.

4.4.5 *Effectiveness of interventions*

By adding the intervention groups to the base model, the analysis explores whether the groups (control, logbook, hazard perception training, education) differ in their collision involvement at 12 months post-test, after considering and controlling for the variables in the base model. The purpose of this analysis is to understand the real-world effectiveness of the interventions regardless of the level of engagement with the intervention for each participant. Put simply, it estimates what happens to collision risk at the population level, if such interventions are offered on a voluntary basis. The population in this case is all learner drivers aged 17-24 in Great Britain, although the previously noted caution needed when generalising from the sample in the study (which is more female than the population and may differ in other ways) is noted.

The variable 'group' was added to the base model as a categorical variable and the coefficients are presented in Table 4-11. One of the groups was arbitrarily chosen as the baseline when the project team was blinded for the analysis (to avoid potential bias in model interpretation). After unblinding, it turned out that the logbook intervention group had been chosen as the baseline.

Table 4-11: Coefficients from the ITT analysis ((*) denotes statistically significant coefficients)

Variable	Coefficient	Standard Error	*p*-value
Intercept	**-3.858**	**0.554**	**<0.01(*)**
Log(Mileage)	**0.297**	**0.064**	**<0.01(*)**
Gender: Male	**-0.084**	**0.092**	**0.36**
Age group: 22-27 years old	**-0.355**	**0.143**	**0.01(*)**
Test pass: Pre-COVID-19 pandemic	**1.624**	**0.745**	**0.03(*)**
Interaction between mileage and test pass	**-0.164**	**0.088**	**0.06**
Group: Education course	-0.114	0.128	0.38
Group: Control	-0.011	0.116	0.92
Group: Hazard perception training	-0.078	0.123	0.52

The analysis showed that the groups did not differ significantly on self-reported collisions in the first 12 months of driving post-test. The response variable for the base model and the ITT analysis was also modelled as a binary outcome (no collisions or one or more collisions) using a logistic regression model. However, the results remained unchanged.

The ITT analysis suggests that if the logbook, hazard perception training, or education interventions in this study are offered on a voluntary basis, with incentives that match what would be feasible in live roll-out but do not excessively incentivise engagement, there is likely to be no safety benefit in terms of collision reduction at the population level (the population being all learner drivers aged 17-24). This is highly likely to be the case even if the interventions are found to be effective in changing surrogate measures of risk for the small proportion of people who do engage. Put simply, engagement levels are so low under voluntary take-up (from 3% to 11%) that any impact on safety is likely to be too small to detect at the population level.

4.5 Per protocol (PP) analysis

This section first presents a background to the per-protocol analysis and the rationale for using it. It then presents an analysis of the characteristics of people engaging with the interventions, to check for any bias that might need to be corrected when comparing surrogate measures. Finally, it presents the analysis, checking for the effectiveness of the interventions in changing the surrogate measures.

The key insights from this analysis are as listed below, separately for the logbook, hazard perception training, and education interventions.

For the logbook intervention:

- Only 3% of participants assigned to the logbook group engaged with the intervention.
- The 'type' of people who engaged with the logbook intervention were more likely to be younger, female and from a 'less deprived' area than those who didn't engage. These differences were accounted for in the analysis.
- In comparison with the control group, those who engaged with the logbook intervention:
    - Were more likely to learn to drive with passengers in the car

  The safety impact that would being expected to result from this change is unclear.
- No other differences were found between those who engaged with the logbook intervention and the control group.

For the hazard perception training intervention:

- Only 11% of participants assigned to the hazard perception training group engaged with the intervention.
- The 'type' of people who engaged with the hazard perception training intervention differed from those who did engage on five characteristics - age, education, social deprivation index, qualifications held and qualifications working towards. These differences were accounted for in the analysis.
- In comparison with the control group, those who engaged with the hazard perception training intervention:

    - Reported a lower frequency of driving above the speed limit in their first three months of post-test driving.
    - Took fewer attempts to pass the driving theory test (suggesting higher hazard perception skill).

  Both of these changes would be expected to have a safety benefit.

- No other differences were found between those who engaged with the education interventions and the control group.

For the education intervention:

- Only 5% of participants assigned to the education group engaged with the intervention.
- The 'type' of people who engaged with the education intervention were more likely to be older and have attained a first degree level qualification than those who did not engage.
- In comparison with the control group, those who engaged with the education intervention:
    - Reported less of their learning to drive as being in the dark.
    - Reported more of their learning to drive as being on dual carriageways.

  The safety impact that would be expected to result from these changes is unclear.

- No other differences were found between those who engaged with the education interventions and the control group.

4.5.1 Background

The per-protocol (PP) analysis looked at those participants in the treatment groups who engaged to some degree with the intervention they were offered. The analysis compared people who actually engaged in some way with their interventions (either the logbook, the hazard perception training, or the education course) with control participants, on a set of surrogate measures of safety (see section 3.2.3 for details). The impact on number of collisions was not explored in the per-protocol analysis as in the ITT analysis due to the small sample size of participants in this analysis. A small sample like this cannot support a robust analysis if using a measure such as collisions as the outcome variable; the low baseline and higher variability of the collisions measure (compared with the surrogate measures) means that much higher sample sizes are needed to support a robust analysis.

Note that in this section, the sample used is not the same as the trial sample used in the previous ITT analysis. Data on whether participants engaged (or did not engage) with the interventions were available (from delivery partners) regardless of whether they passed their test or completed any data collection surveys. Therefore, for the matching process described in section 4.5.2, all participants recruited for the learner arm were initially included. For the analysis of effectiveness in section 4.5.3, the sample used depends on the measures being analysed – this is specified in the text where relevant.

4.5.2 Characteristics of engagers and correcting for engager bias

An important part of the per-protocol analysis was to ensure as far as possible that when comparing the treatment engagers with the control participants, this would be comparing 'like for like' in every way possible except the receiving of the intervention. It would not be possible to simply compare engagers in a treatment group with the entire control group; this could well lead to bias as it would be comparing one 'type' of person in a treatment group (the type of person who engages) with at least two 'types' in the control group (people who do, and do not, engage).

Therefore, checks were made for engager bias before running the analysis to assess whether the measured characteristics of people who engaged with the interventions were different to those of people who did not, and to determine whether it was necessary to sample the control group to ensure that the characteristics of the control group participants matched those of the engagers in the respective intervention group. Checking for engager bias was done by comparing engagers and non-engagers (using statistical tests) for each intervention on a number of different characteristics such as age, gender, education status, and social deprivation. In addition, a 30-item personality inventory was used in the registration survey to score people on the 'Big Five' personality factors (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism). Comparisons performed on these personality variables revealed some differences between engagers and non-engagers for each group, but with negligible effect sizes. Therefore, personality variables were not used when sampling the control group to find a matching sample. While a range of measures were used for this checking procedure, it remains possible that the engagers and non-engagers might have differed on other unknown variables.

For two of the interventions (hazard perception training and education course), engagers were categorised as 'engagers' and 'high engagers' (see sections 4.5.2.2 and 4.5.2.3). Checks for engager bias and subsequent sampling of the control group were only done once for these groups using the entire engager sample.

In all cases the threshold for what constituted being an 'engager' was relatively low. This needs to be considered when interpreting the results; because engagement 'dose' or amount of contact with the interventions was low even in those people who did engage, it may be that the effects seen were limited by this.

4.5.2.1 Logbook intervention

An engager with the logbook intervention was defined as someone who used the app and recorded at least one session. The information used to identify engagers came from DIA, who provided app data for each participant. Given the very low numbers of people engaging with the app, a very modest criterion had to be adopted. Anyone who had completed at least one recorded journey using the app was classed as having engaged. Table 4-12 shows the breakdown of participants in the logbook group by engagement. Of the 4,438 participants in the logbook group, only 3% (121 participants) were identified as engagers. That is to say – of the people given the opportunity to use the app, only 3% recorded at least one session. The highest number of sessions recorded by a participant was 48.

Table 4-12: Summary of engagement for logbook group

	Number of participants	% of participants in logbook group
Total in logbook group	4,438	100%
Total who recorded at least one session (engagers)	**121**	**3%**

The characteristics of the 121 engagers were compared with those of the 4,317 non-engagers to determine if the 'type' of person who engaged with the mentoring agreement differed from the 'type' who did not. There were three variables where differences were identified: Age (at registration), Gender and SDI (shown in Figure 4-19, Figure 4-20 and Figure 4-21 respectively). Charts showing the full results of all the comparisons carried out are presented in Appendix D.

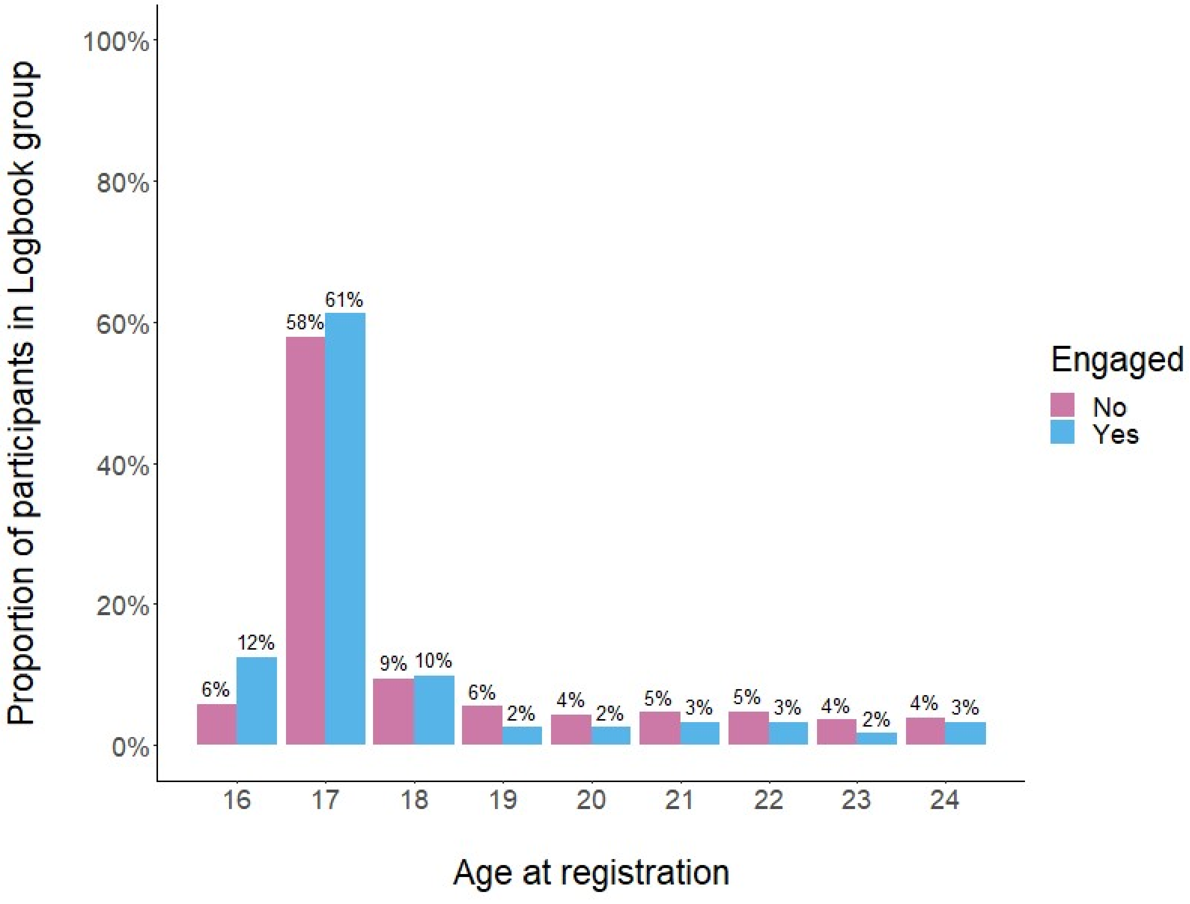


Figure 4-19: Proportion of engagers and non-engagers in the logbook group by age (121 engagers, 4,317 non-engagers)

Figure 4-19 shows that the proportion of participants in the 16 to 18-year-old categories is higher for engagers and the proportion in the 19 to 24-year-old categories is higher for non-engagers. This suggests that younger people may be more likely to engage with the logbook intervention. The mean age (at registration) for non-engagers was 18.2 and for engagers it was slightly lower at 17.7. Statistical tests showed that there was a significant difference in the distribution of engagers and non-engagers by age at test pass (p=0.01) with a small effect size of 0.05.

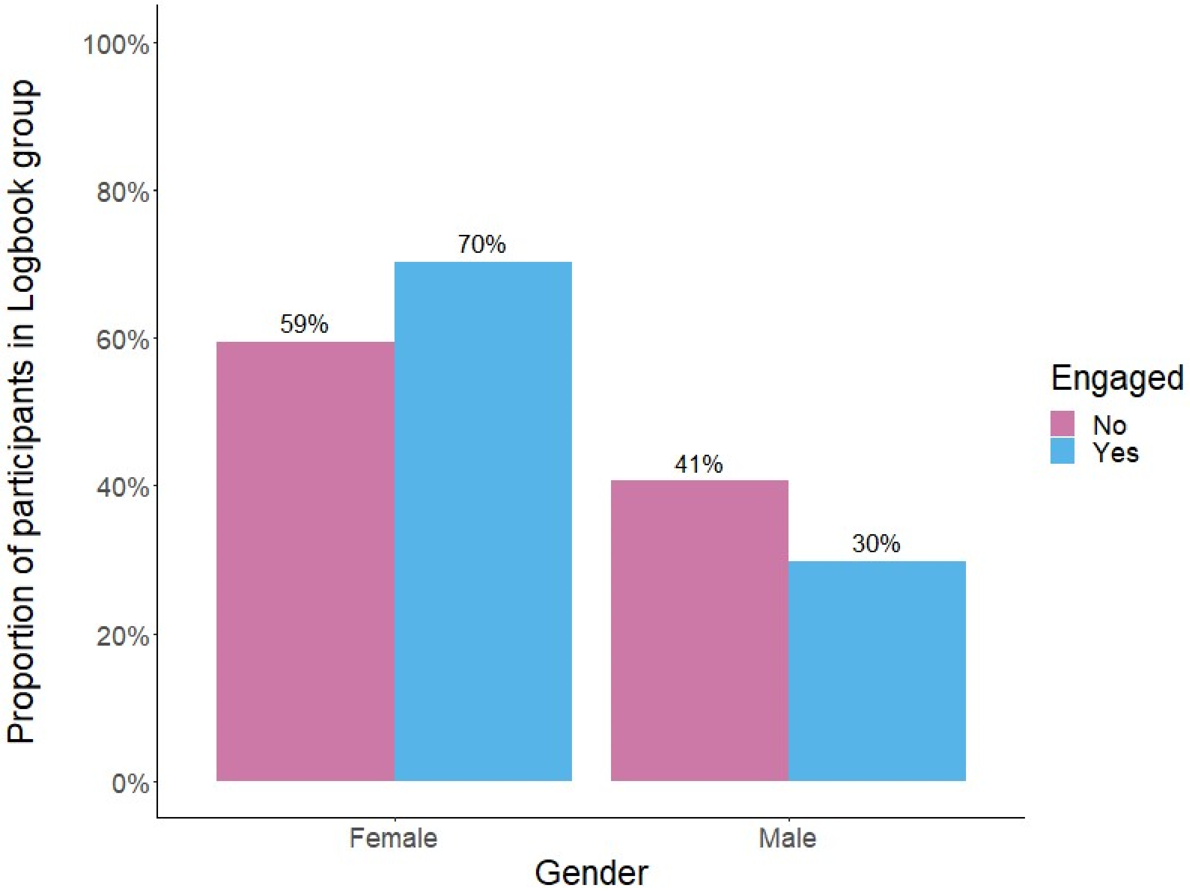


Figure 4-20: Proportion of engagers and non-engagers in the logbook group by gender (121 engagers, 4,317 non-engagers)

The results in Figure 4-20 show that engagers with the logbook intervention were more likely to be female than male; 70% of engagers were female, compared with 59% of non-engagers. Statistical tests showed a significant difference in the distribution of engagers and non-engagers by gender (p=0.02) with a very small effect size of 0.03.

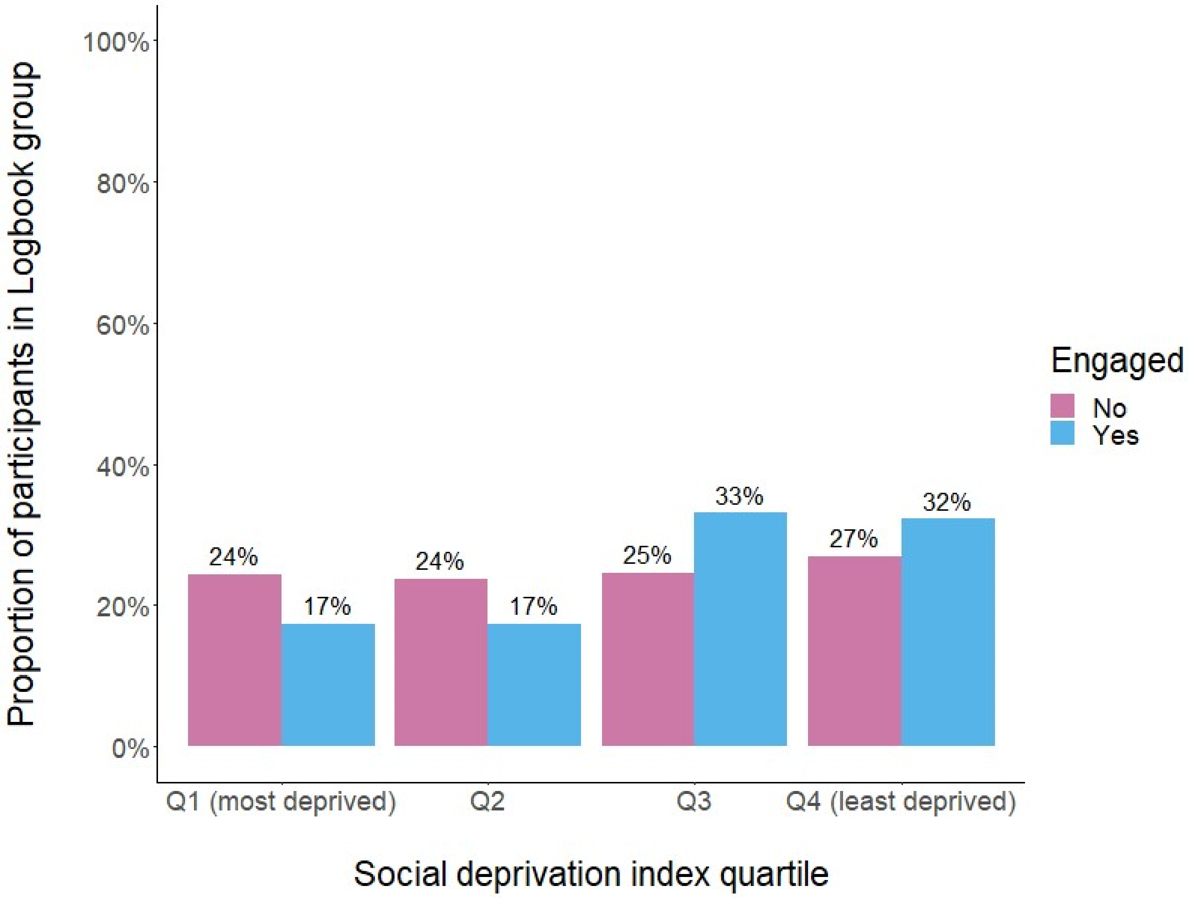


Figure 4-21: Proportion of engagers and non-engagers in the logbook group by social deprivation index (SDI) quartile (121 engagers, 4,317 non-engagers)

The results in Figure 4-21 show that logbook engagers were more likely to be in Q3 and Q4 (least deprived) compared with non-engagers; 65% of engagers were in Q3 or Q4, compared with 51% of non-engagers. Statistical tests showed a significant difference in the distribution of engagers and non-engagers by SDI quartile (p=0.03) with a small effect size of 0.05.

The differences identified between engager and non-engager characteristics, despite the small effect sizes, meant that it was necessary to correct for engager bias in the control group, in order for comparisons between engagers and control participants to be robust. Therefore, a nested case control method was used to create a sample of control group participants with characteristics matching those of the participants who had engaged with the logbook intervention. For details on this process, see Appendix E.1.

Table 4-13 shows the number of logbook engagers and participants from the control group sample who completed each survey, and the total number of logbook engagers and control group sample participants included in the per-protocol analysis.

Table 4-13: Summary of participants included in per-protocol analysis for the logbook intervention

Survey	Number of logbook engagers	Number of control group participants
Test pass	84	410
3-month	79	347
6-month	70	316
12-month	66	275
Total filling out all surveys	63	245
Total included in PP analysis - answered at least one survey	**85**	**440**

Note that the totals are lower than 121 and 961 as there were 36 logbook engagers and 521 participants in the control group sample who did not answer any surveys (apart from the registration survey). Many of these participants not answering any surveys did not pass their test during the trial period.

4.5.2.2 Hazard perception training intervention

An engager with the hazard perception training intervention was defined as someone who completed at least one module. A 'high engager' was defined as someone who completed at least two modules. Table 4-14 shows the breakdown of participants in the hazard perception training group by level of engagement. Of the 3,851 participants in the hazard perception training group, 412 were identified as engagers and 131 as high engagers. Forty-two participants completed all three modules.

Table 4-14: Summary of engagement for hazard perception training group

	Number of participants	% of participants in hazard perception training group
Total in hazard perception training group	3,851	100%
Total who completed at least one module (engagers)	**412**	**11%**
Total who completed at least two modules (high engagers)	**131**	**3%**

The characteristics of the 412 engagers were compared with those of the 3,439 non-engagers. There were five variables where differences were identified: age, education status, SDI quartile, highest qualification achieved and qualification hoping to get (shown in Figure 4-22, Figure 4-23, Figure 4-24, Figure 4-25 and Figure 4-26 respectively). Charts showing the results for the other variables examined are presented in Appendix D.

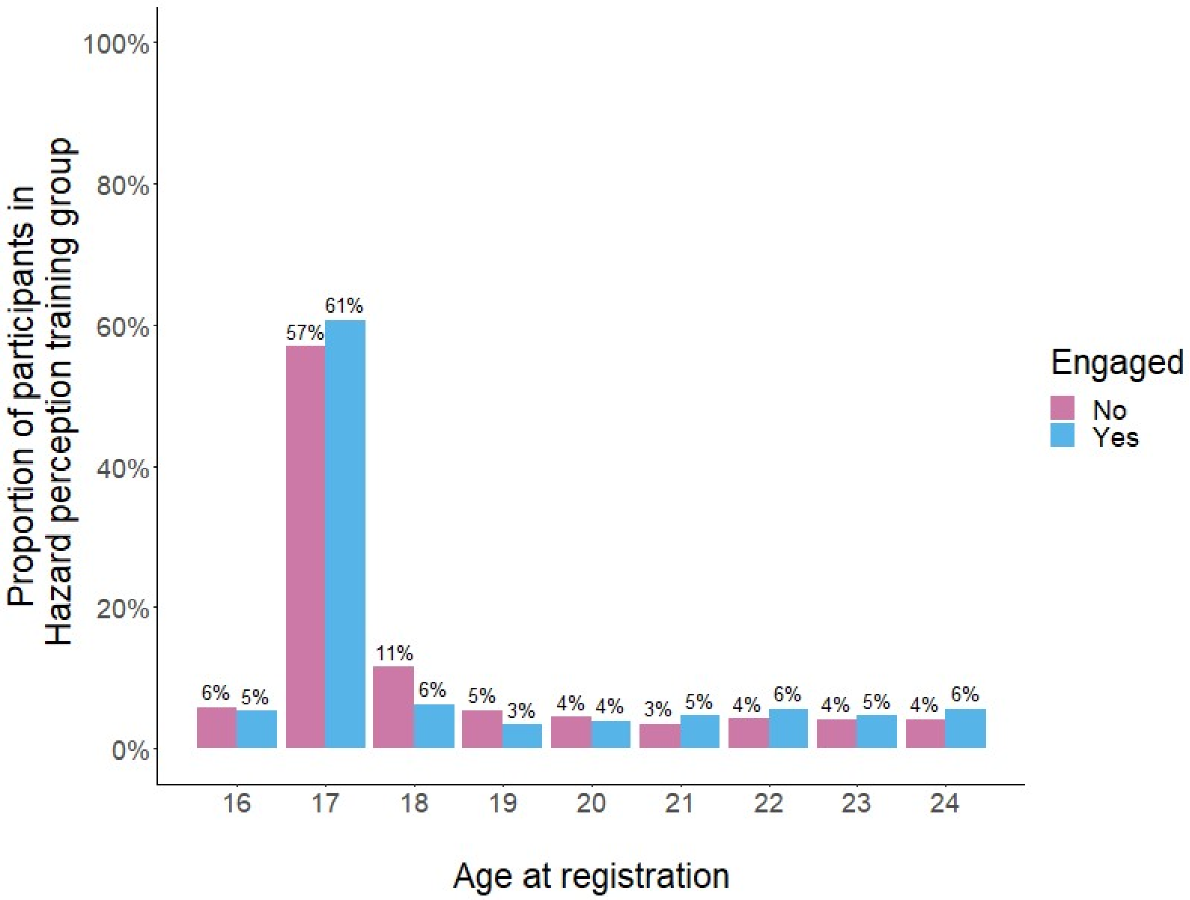


Figure 4-22: Proportion of engagers and non-engagers in the hazard perception training group by age (412 engagers and 3,439 non-engagers)

Comparing age distributions, Figure 4-22 shows that the most notable difference between engagers and non-engagers is the proportion of 18-year-olds; 11% of non-engagers were 18 at registration compared with 6% of engagers. Whilst there are differences by individual age categories, the mean ages for the two groups are very similar at 18.2 for non-engagers and 18.3 for engagers. Statistical tests showed that there was a significant difference in the distribution of engagers and non-engagers by age at registration (*p*=0.02) with a small effect size of 0.07.

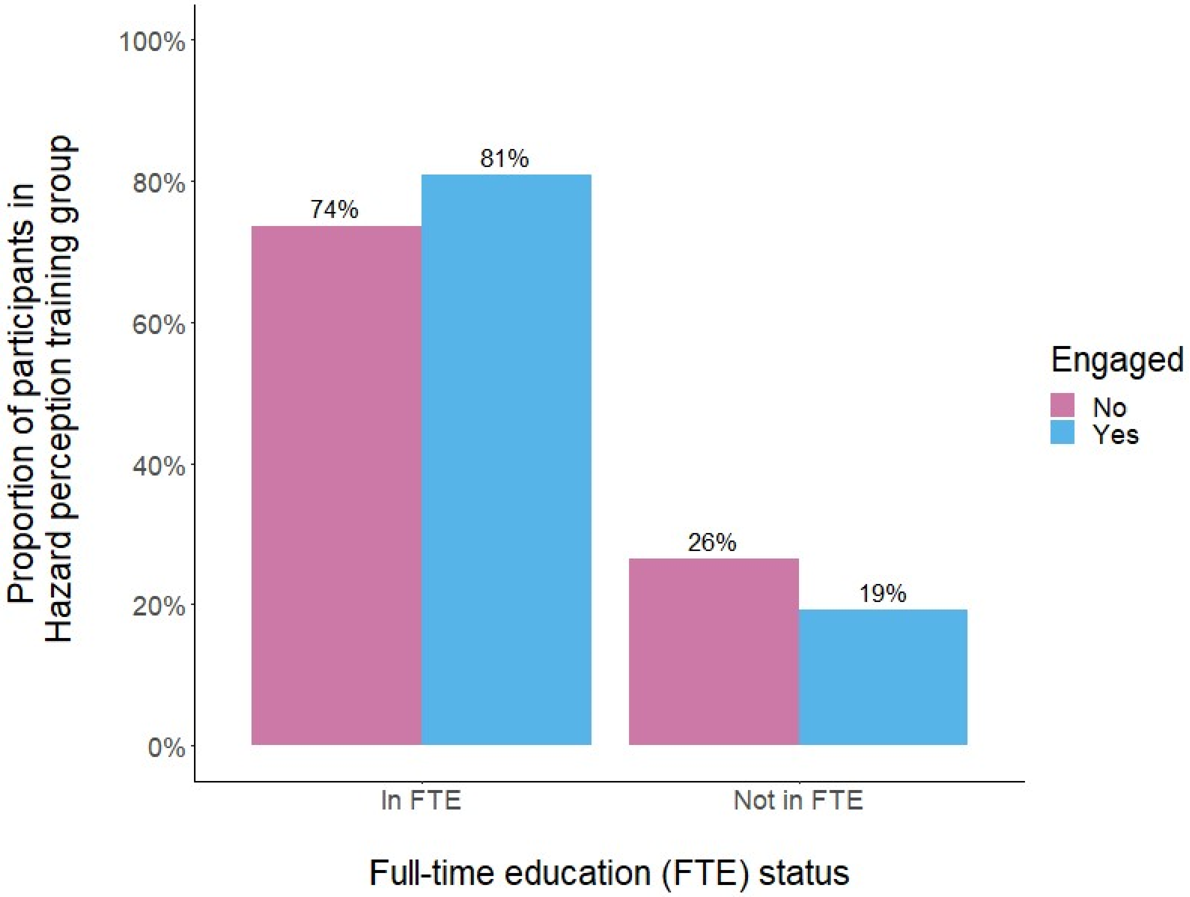


Figure 4-23: Proportion of engagers and non-engagers in the hazard perception training group by education status (412 engagers and 3,439 non-engagers)

The results in Figure 4-23 show that engagers with the hazard perception training intervention were more likely to be in full time education; 81% of engagers were in full time education, compared with 74% of non-engagers. Statistical tests showed a significant difference in the distribution of engagers and non-engagers by education status ($p<0.01$) with a small effect size of 0.05.

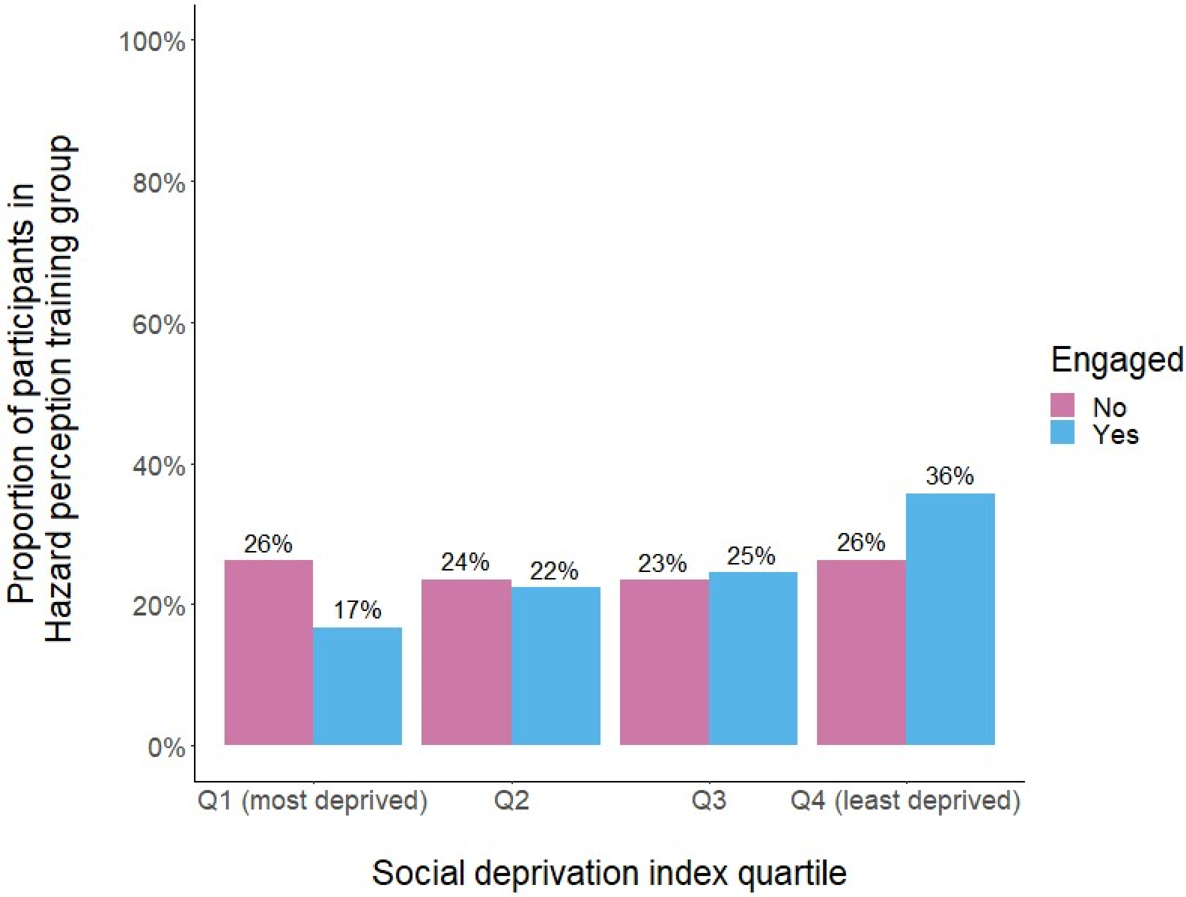


Figure 4-24: Proportion of engagers and non-engagers in the hazard perception training group by SDI quartile (412 engagers and 3,439 non-engagers)

Figure 4-24 shows that there was a higher proportion of engagers (36%) in Q4 (least deprived) compared with non-engagers (26%). There was also a higher proportion of non-engagers (26%) in Q1 (most deprived) compared with engagers (17%). Statistical tests showed a significant difference in the distribution of engagers and non-engagers by SDI quartile ($p<0.01$) with a small effect size of 0.08, with those least deprived most likely to engage.

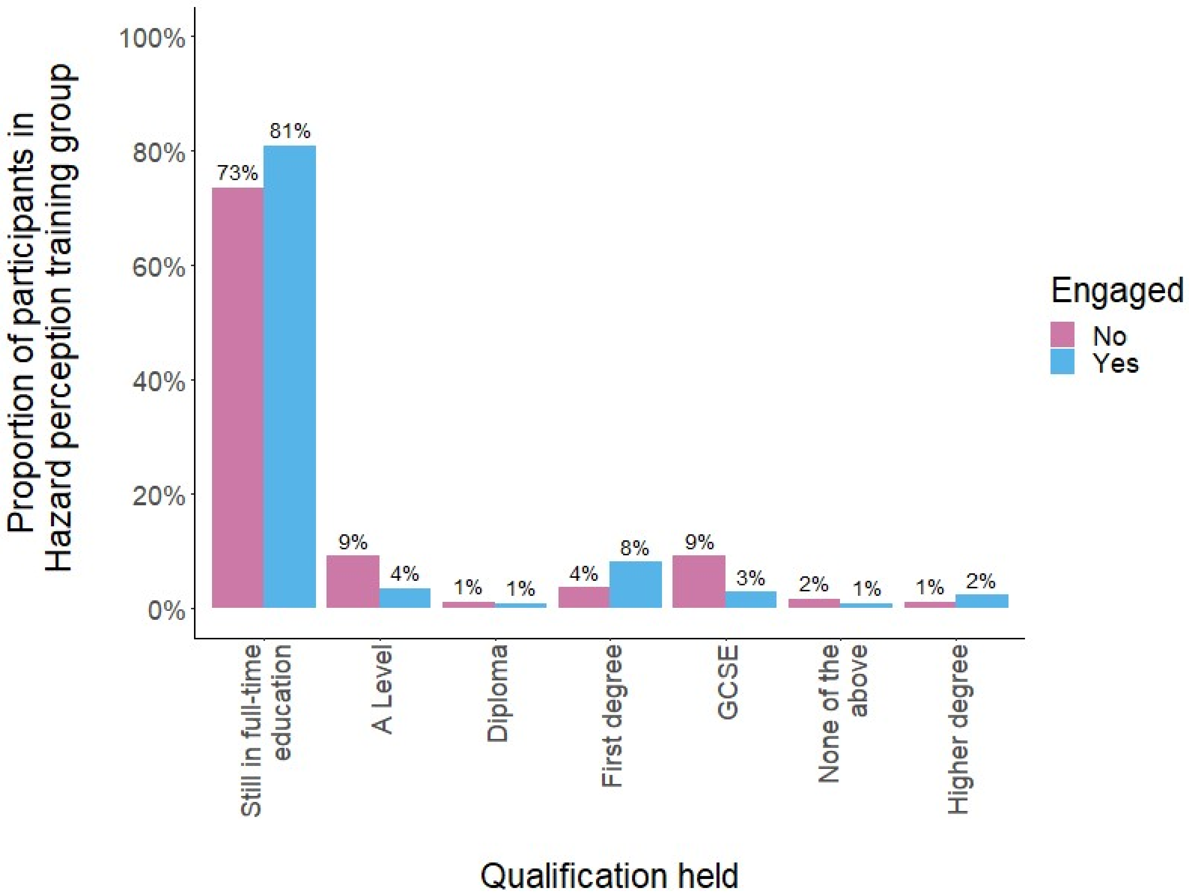


Figure 4-25: Proportion of engagers and non-engagers in the hazard perception training group by highest education qualification held (412 engagers and 3,439 non-engagers)

Figure 4-25 shows that non-engagers were more likely to hold GCSEs or A levels as their highest qualification; 18% of non-engagers held GCSEs or A levels as their highest qualification compared with 7% of engagers. Engagers were more likely to still be in education or hold their first degree as highest qualification. Statistical tests showed a significant difference in the distribution of engagers and non-engagers by highest qualification held ($p<0.01$) with a small to medium effect size of 0.11.

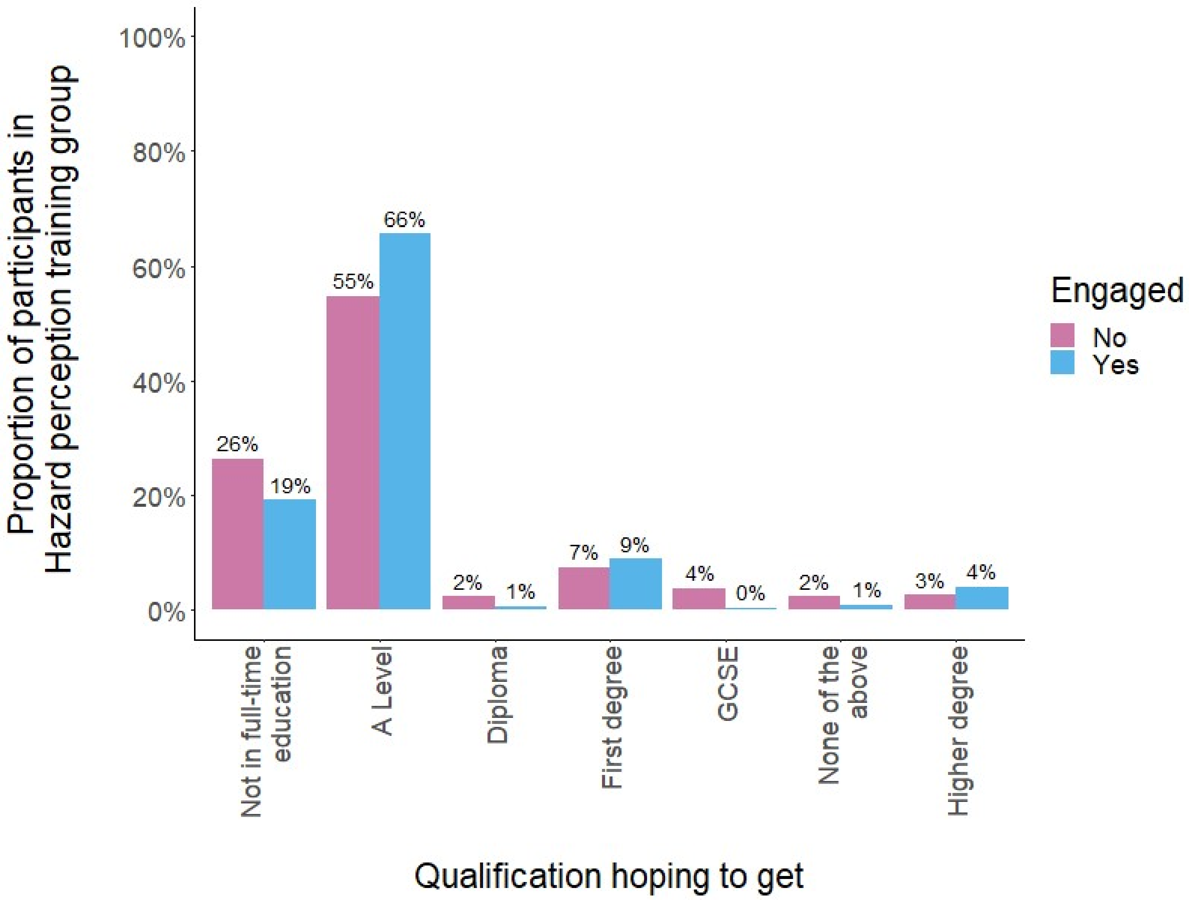


Figure 4-26: Proportion of engagers and non-engagers in the hazard perception training group by education qualification they hope to get (412 engagers and 3,439 non-engagers)

Figure 4-26 shows that a higher proportion of engagers were hoping to get their A-levels; 66% of engagers were in this category compared with 55% of non-engagers. A higher proportion of non-engagers were not in full time education (26%) or were studying for their GCSEs (4%), compared with engagers (19% and <1% respectively). Statistical tests showed a significant difference in the distribution of engagers and non-engagers by the qualification they were hoping to get (p<0.01) with a small effect size of 0.09.

The differences identified between engager and non-engager characteristics, despite the small effect sizes, meant that it was necessary to correct for engager bias in the control group, in order for comparisons between engagers and control participants to be robust. Therefore, a nested case control method was used to create a sample of control group participants with characteristics that matched those of the participants who had engaged with the hazard perception training intervention. For more detail on this process, see Appendix E.2.

Table 4-15 shows the number of hazard perception training engagers and participants from the control group sample who completed each survey, and the total number of hazard perception training engagers and control group sample participants included in the per-protocol analysis. The number of participants included in the ‘number of theory test attempts’ analysis – see Table 4-26 – is slightly more than the total number in Table 4-15 as all test passers were included regardless of whether they had provided any survey data.

Table 4-15: Summary of participants included in per-protocol analysis for the hazard perception training intervention

Survey	Number of hazard perception training engagers (high engagers)	Number of control group sample participants
Test pass	292 (115)	525
3-month	263 (109)	445
6-month	253 (104)	410
12-month	233 (106)	345
Total filling out all surveys	218 (100)	307
Total included in PP analysis - answered at least one survey	**300 (117)**	**549**

Note that the totals are lower than 412 and 1,220 as there were 112 engagers and 671 participants in the control group sample who did not answer any surveys (apart from the registration survey). Many of these participants not answering any surveys did not pass their test during the trial period.

Given that there were still reasonable sample sizes in the survey time periods for the high engagers, it was decided that this group would be analysed separately.

4.5.2.3 Education course intervention

An engager with the education course intervention was defined as someone who attended the course in-person or completed at least one module online. A 'high engager' was defined as someone who attended the course in-person or completed at least three modules online. Table 4-16 shows the breakdown of participants in the education course group by level of engagement. Of the 3,850 participants in the education course group, 181 were identified as engagers and 137 as high engagers; due to the very low sample sizes at each survey time point, the high engager analysis was not undertaken. Of the 181 engagers, 110 did the course in-person, 69 did the online version and two participants did both.

Table 4-16: Summary of engagement for education course group

	Number of participants	% of participants in education course group
Total in education course group	3,850	100%
Total who attended the course in-person or completed at least one module online (engagers)	**181**	**5%**
Total who attended the course in-person or completed at least three modules online (high engagers)	**137**	**4%**

The characteristics of the 181 engagers were compared with those of the 3,669 non-engagers. There were two variables where differences were identified: age and highest qualification held (shown in Figure 4-27 and Figure 4-28 respectively). Charts showing the results for the other variables examined are presented in Appendix D.

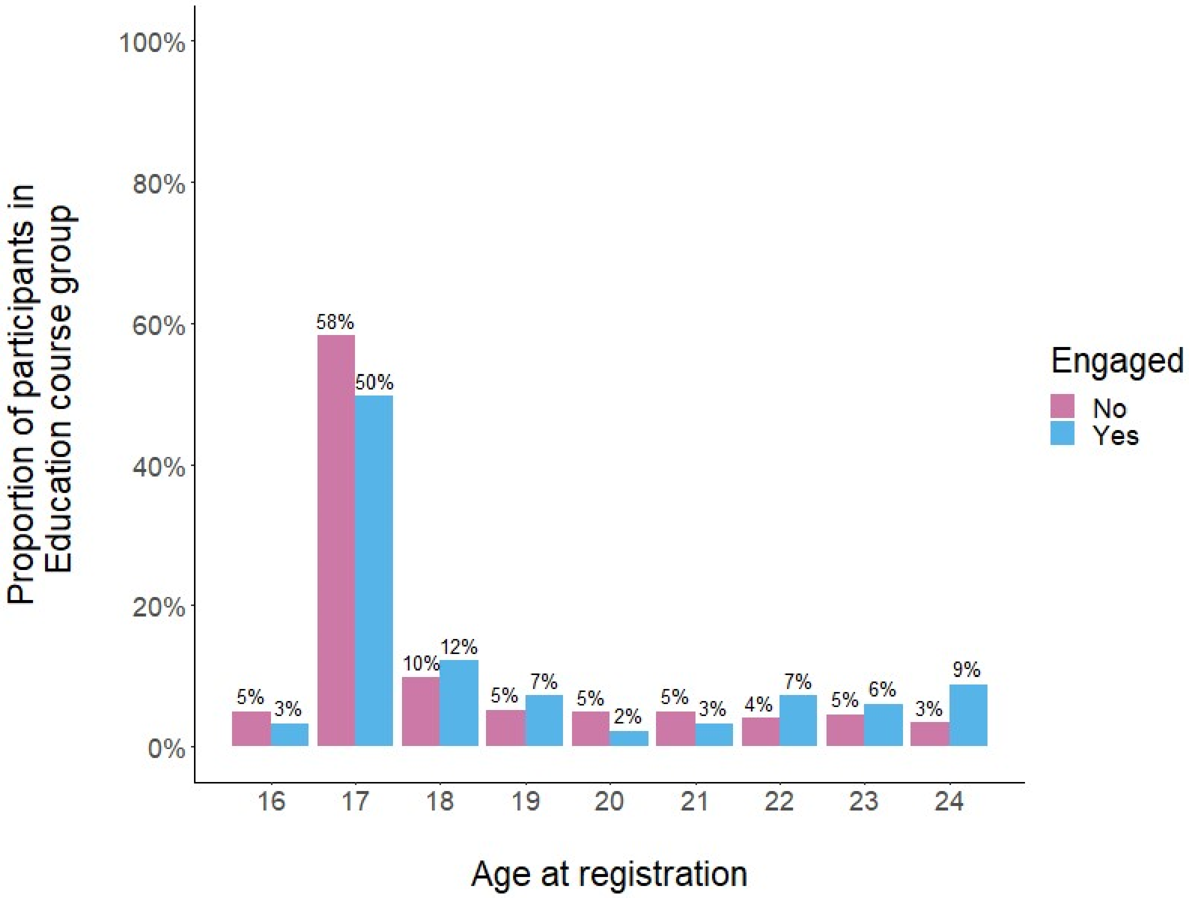


Figure 4-27: Proportion of engagers and non-engagers in the education course group by age (181 engagers and 3,669 non-engagers)

Figure 4-27 shows that the proportion of participants in each age category differs between engagers and non-engagers. Most notably, the proportion of 17-year-olds is higher for non-engagers (58%) than engagers (50%) and the proportion of 24-year-olds is much higher for engagers (9%) than non-engagers (3%). Statistical tests showed a significant difference in the distribution of engagers and non-engagers by age (p<0.01) with a small effect size of 0.09.

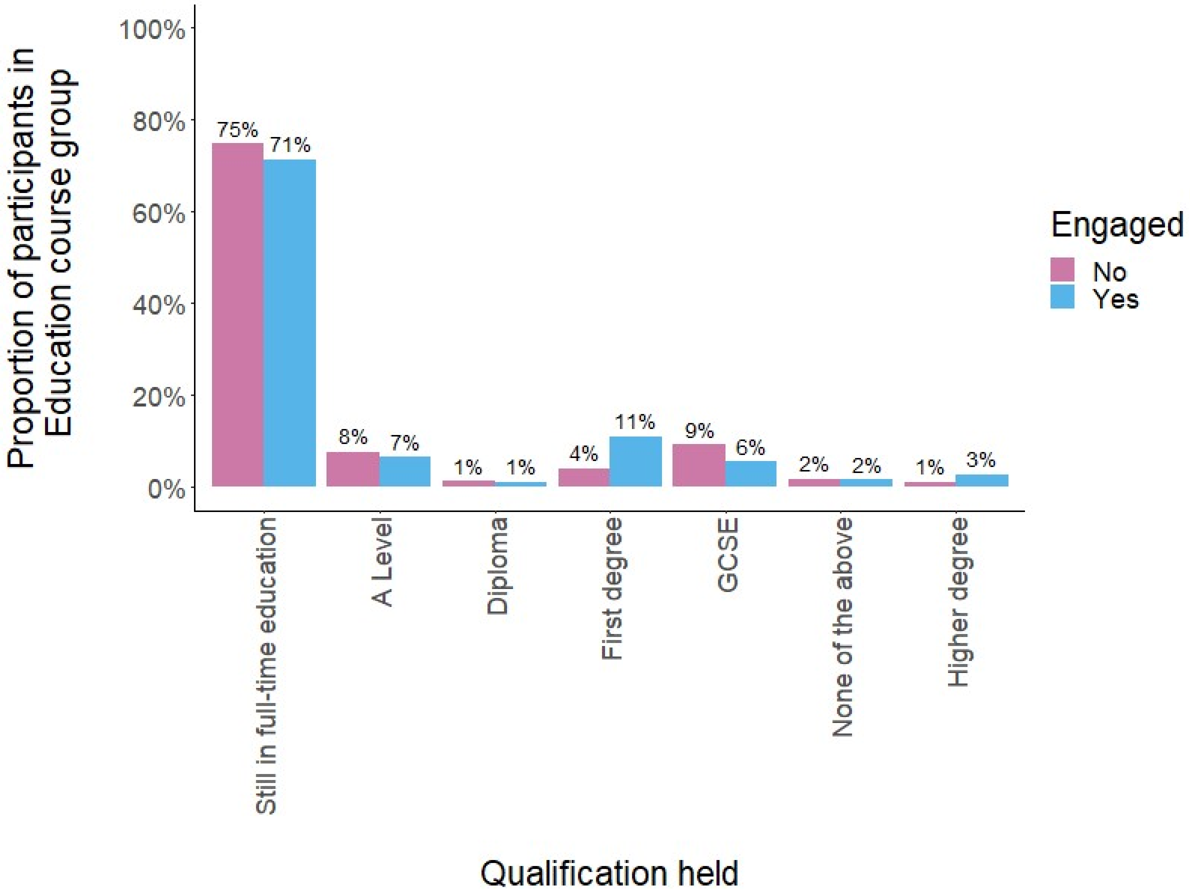


Figure 4-28: Proportion of engagers and non-engagers in the education course group by highest education qualification held (181 engagers and 3,669 non-engagers)

Figure 4-28 shows that engagers (11%) were more likely to hold a first degree as their highest qualification than non-engagers (4%). Non-engagers were more likely to hold GCSEs as their highest qualification (9% compared with 6% for engagers) or to still be in full-time education (75% compared with 71% for engagers). Statistical tests showed a significant difference in the distribution of engagers and non-engagers by highest education qualification held (p<0.01) with a small effect size of 0.07.

The differences identified between engager and non-engager characteristics, despite the small effect sizes, meant that it was necessary to correct for engager bias in the control group, in order for comparisons between engagers and control participants to be robust. Therefore, a nested case control method was used to create a sample of control group participants with characteristics that matched those of the participants who had engaged with the logbook intervention. For more detail on this process, see Appendix E.3.

Table 4-17 shows the number of education course engagers and participants from the control group sample who completed each survey, and the total number of education course engagers and control group sample participants included in the per-protocol analysis.

Table 4-17: Summary of participants included in per-protocol analysis for the education intervention

Survey	Number of education course engagers (high engagers)	Number of control group sample participants
Test pass	79 (57)	416
3-month	71 (51)	356
6-month	68 (49)	331
12-month	61 (45)	277
Total filling out all surveys	59 (43)	249
Total included in PP analysis - answered at least one survey	**82 (60)**	**440**

Note that the totals are lower than 181 and 1,043 as there were 99 engagers and 603 participants in the control group sample who did not answer any surveys (apart from the registration survey). Many of these participants not answering any surveys did not pass their test during the trial period.

Given the very low sample sizes at the survey time points, it was decided that the 'high engager' group would not be analysed separately.

4.5.3 *Effectiveness of interventions*

This section presents the findings on the effectiveness of the three interventions in changing the relevant surrogate outcomes defined. In other words, for people who engage, what does the intervention change (relative to matched control group participants)?

The engagers with each intervention were compared with the relevant control group participants over each of the reporting periods to assess what difference the intervention made to a variety of measures. As noted in section 2.2.1, some measures assessed were intervention-specific (e.g., number of limits set on driving in certain conditions or driving style) and some were general surrogate safety measures (e.g. number of near misses).

The number of engagers and control group participants in each comparison varies for each reporting period because some participants did not answer all surveys. However, the results for the analysis of variance (ANOVA) statistical tests (which are performed to look at differences by time, group and the interaction between time and group) in this section only use the data from participants filling out all surveys, as they require no missing entries. The number of these participants is given in Table 4-13, Table 4-15 and Table 4-17 for each of the three intervention groups.

4.5.3.1 *Logbook intervention*

This section presents the results of analysis which seeks to answer the question: for people who engage, what does the logbook intervention change? Logbook engagers and control group participants were compared using the following measures to try to answer this question.

- Self-reported near misses – testing whether there were fewer for those who engaged with the logbook intervention
- Self-reported Driving Events scale score – tested whether this was lower (indicating lower frequency of unsafe driving events) for those who engaged
- Self-reported frequency of breaking the speed limit when driving – tested whether this was less frequent for those who engaged
- Self-reported hours spent driving with an ADI before passing the practical test – tested whether this was more for those who engaged
- Self-reported hours spent driving with another supervising driver (not an ADI) before passing the practical test - tested whether this was more for those who engaged
- Self-reported proportion of learning mileage spent driving on different road types, testing if learning was more spread between road types for logbook engagers.

All comparisons in this section have been made between 85 participants who engaged with the logbook intervention and the 440 matched control group participants; however, numbers of participants vary by reporting period. See section 4.5.2.1 for the definition of a logbook intervention engager, and a summary of the number of participants in each group included in the analysis of each reporting period.

Near misses

At each time point, participants were asked how many 'near misses' they had experienced during that reporting period. A near miss was described as the impression of only just avoiding an accident.

Figure 4-29 shows the distribution of number of near misses (corrected for mileage) reported by participants who engaged with the logbook intervention and control group participants, during each of the reporting periods (months 1-3, months 4-6 and months 7-12). This is presented as a boxplot which displays the inter-quartile range (25^{th} and 75^{th} percentiles) as the top and bottom of the box, and the median (50^{th} percentile or 'average') as the horizontal line in the box. Values outside of the 95% confidence interval are shown as individual data points. The mean number of near misses per mile reported for both groups across the three periods is shown in Table 4-18.

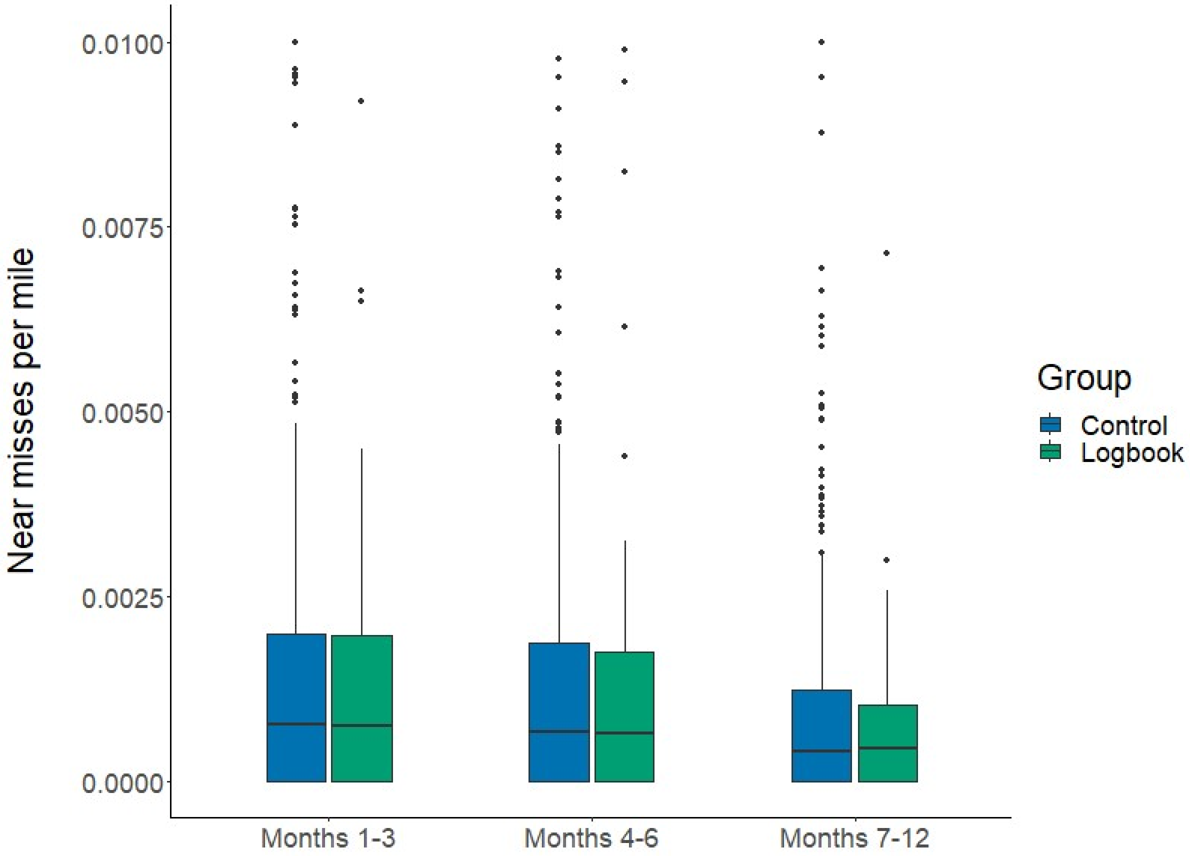


Figure 4-29: Distribution of number of reported near misses per mile in each post-test reporting period for logbook engagers and control group participants (Months 1 to 3: 79 engagers, 347 control group. Months 4 to 6: 70 engagers, 316 control group. Months 7 to 12: 66 engagers, 275 control group).

Table 4-18: Reported number of near misses per mile in each post-test reporting period for logbook engagers and control group participants (Months 1 to 3: 79 engagers, 347 control group. Months 4 to 6: 70 engagers, 316 control group. Months 7 to 12: 66 engagers, 275 control group).

	Mean number of near misses per mile driven		
Group	Months 1-3	Months 4-6	Months 7-12
Control group participants	0.001	0.001	0.0005
Logbook engagers	0.001	0.0009	0.0005

Engagers with the logbook intervention and control group participants all reported low numbers of near misses per mile in all reporting periods. The rates per mile are shown in Table 4-18. For context, in months 1-3 and 4-6 people were reporting around one near miss for every 1,000 miles of driving, with this dropping to around one every 2,000 miles of driving in months 7-12. The number of near misses per mile was very similar between groups across all reporting periods. Both groups reported a smaller number of near misses per mile in months 7-12 compared with months 1-3 and months 4-6.

Statistical tests showed no significant differences between the two groups (p=0.63). The differences over time were also not significant for either group (p=0.41). The interaction effect was not significant (p=0.22), meaning that the effect of time was the same for both groups.

Driving events score

In each post-test survey, participants were asked about how often each of six driving events had occurred whilst they were driving during that reporting period:

- Braking sharply to avoid a collision with the vehicle ahead because it had slowed
- Pulling out to overtake or turn right not noticing another vehicle in their 'blind spot'
- Failing to notice someone waiting at a pedestrian crossing
- When cornering, finding they were traveling too fast to negotiate the bend safely and having to brake
- Failing to give way when entering a roundabout to a vehicle already on the roundabout
- Braking or swerving suddenly to avoid an accident

Participants indicated how often each of the above events had occurred using a six-point scale ranging from ‘never’ (a score of 1) to ‘nearly all the time’ (a score of 6). The sum of these scores across all six events was then calculated to give an overall ‘driving events score’ A lower driving events score indicates fewer occurrences of the driving events and therefore theoretically safer driving.

Figure 4-30 shows the distribution of driving events scores for each reporting period for participants who engaged with the logbook intervention and participants in the control group. The corresponding means are shown in Table 4-19.

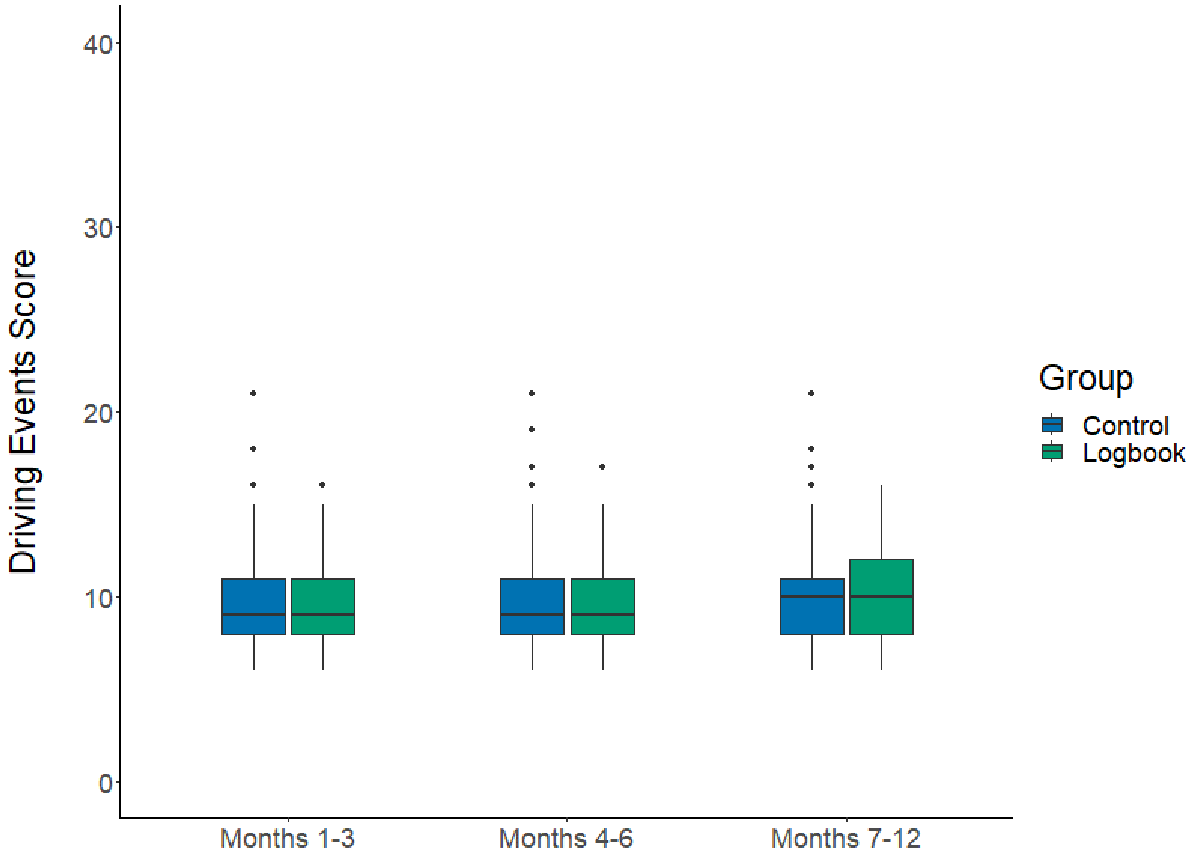


Figure 4-30: Distribution of driving events scores for each post-test reporting period for logbook engagers and control group participants (Months 1 to 3: 79 engagers, 347 control group. Months 4 to 6: 70 engagers, 316 control group. Months 7 to 12: 66 engagers, 275 control group).

Table 4-19: Mean driving event score in each post-test reporting period for logbook engagers and control group participants (Months 1 to 3: 79 engagers, 347 control group. Months 4 to 6: 70 engagers, 316 control group. Months 7 to 12: 66 engagers, 275 control group).

	Mean driving event score		
Group	Months 1-3	Months 4-6	Months 7-12
Control group participants	9.60	9.65	9.96
Logbook engagers	9.57	9.82	9.78

For all reporting periods, the distributions of driving events scores were very similar for logbook engagers and control group participants, with the mean scores between 9 and 10 across all reporting periods for both groups.

There were no clear patterns seen in the mean driving events score across reporting periods and all differences between groups and periods were very small. In months 1-3 the groups had almost identical scores. In months 4-6 logbook participants had slightly higher scores, and in months 7-12, they had slightly lower ones.

Statistical tests showed no significant differences between groups (*p*=0.62). The difference over time was significant for both groups combined (*p*=0.02) with an effect size of close to zero . The interaction between group and time was not significant (*p*=0.44) meaning that the effect of time was the same for both groups.

Exceeding speed limit

In each post-test survey, participants where asked how often they thought they had exceeded the speed limit whilst driving during the relevant reporting period. Participants indicated their frequency of exceeding speed limits using a number from 0 (never) to 100 (all the time). The distributions of these scores for logbook engagers and control group participants is shown in Figure 4-31 and the mean frequency scores are shown in Table 4-20.

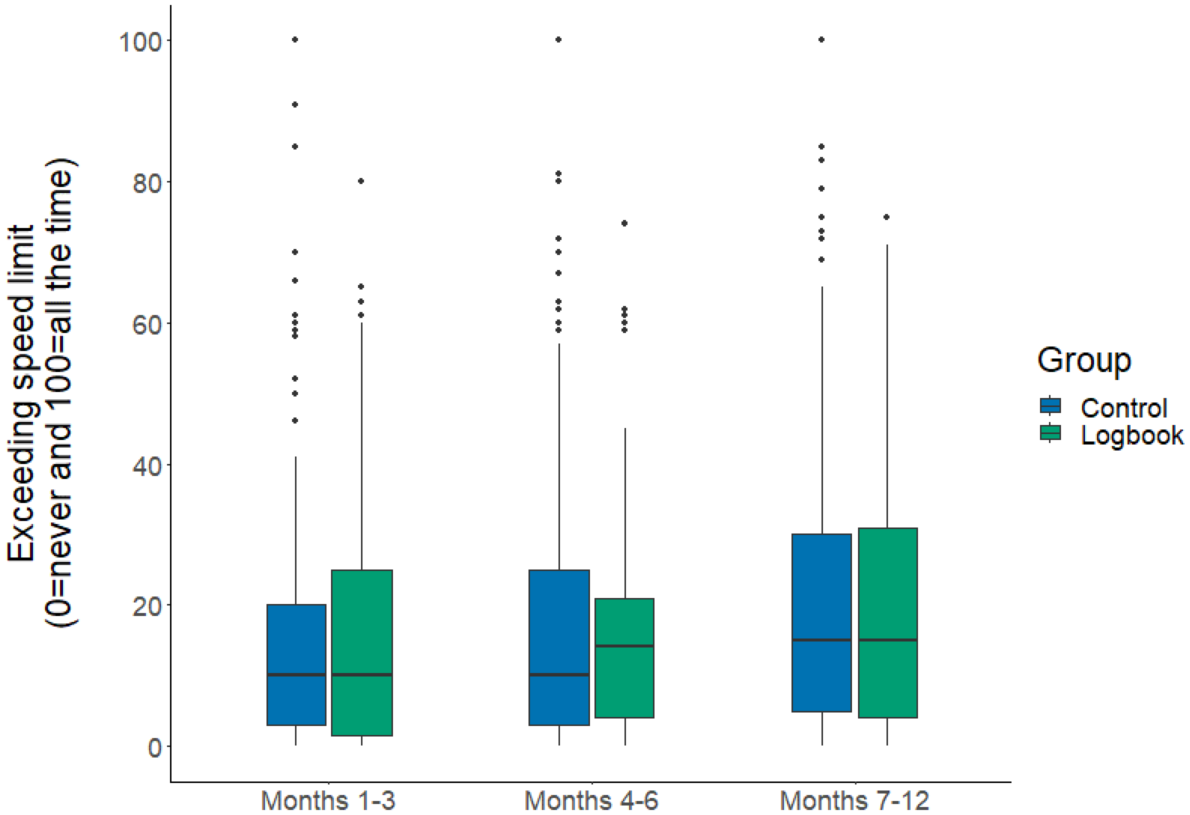


Figure 4-31: Distribution of how often participants thought they exceeded the speed limit for logbook engagers and the corresponding control group sample (Months 1 to 3: 79 engagers, 347 control group. Months 4 to 6: 70 engagers, 316 control group. Months 7 to 12: 66 engagers, 275 control group).

Table 4-20: Mean exceeding speed limit frequency score in each post-test reporting period for logbook engagers and control group participants (Months 1 to 3: 79 engagers, 347 control group. Months 4 to 6: 70 engagers, 316 control group. Months 7 to 12: 66 engagers, 275 control group).

	Mean frequency score		
Group	Months 1-3	Months 4-6	Months 7-12
Control group participants	14.2	17.0	20.4
Logbook engagers	17.2	16.7	20.3

The distributions of exceeding speed limit frequency scores are generally similar between groups, although logbook engagers reported slightly higher mean frequency scores than control group participants in months 1-3. For both groups, the frequency of exceeding speed limits increased appreciably between months 1-3 and months 7-12.

Statistical tests showed there was no significant difference between groups (p=0.20). The difference over time was significant for both groups combined (p<0.01) with a small effect size of 0.01; however, the interaction between group and time was not significant (p=0.45), meaning the effect of time was the same for both groups.

Hours spent learning with an ADI and other supervising drivers

In the test pass survey, participants were asked how many hours they had spent driving with an ADI and with another supervising driver before passing the practical test. The mean number of hours with an ADI and another supervising driver for the logbook engagers and control group participants is shown in Table 4-21.

Table 4-21: Mean hours reported spent learning with an ADI and with another supervising driver for logbook engagers and control group participants (84 engagers, 410 control group)

Group	Mean hours with an ADI	Mean hours with another supervising driver	Mean total hours of practice during learning
Control group participants	41.5	34.8	76.3
Logbook engagers	44.1	37.7	81.8

Table 4-21 shows that the logbook engagers spent more time learning to drive with an ADI and with other supervising drivers, on average, compared with the control group participants. However, these differences were not statistically significant. Statistical tests showed no differences between groups in the number of hours spent with an ADI (p=0.49) or with another supervising driver (p=0.28).

Proportion of mileage learning in different conditions

In the test pass survey, participants were asked about the number of miles spent driving in different conditions before passing the practical test, including on different road types, with additional passengers (not including the supervising driver), in the dark, and in wet conditions. The proportion spent learning in each condition has been averaged across the participants in each group.

Figure 4-32 shows the proportion of learning miles reported to be spent driving in different conditions for logbook engagers and control group participants.

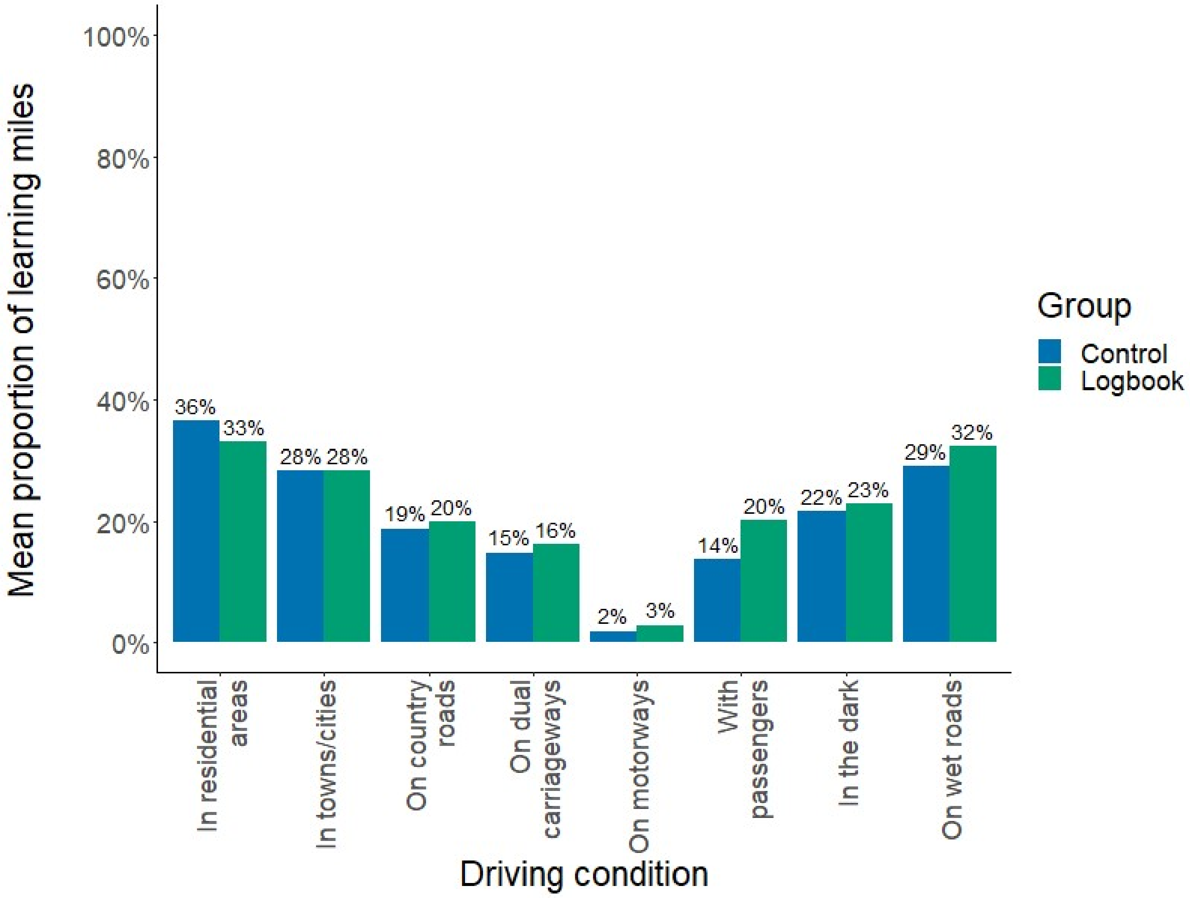


Figure 4-32: Proportion of total learning miles reported to be spent driving in different conditions for logbook engagers and the corresponding control group sample (84 engagers, 410 control group).

Figure 4-32 indicates some differences in the proportion of learning miles spent in different conditions between the logbook engagers and control group participants. The biggest differences can be seen in the proportion of miles spent in residential areas, with passengers and on wet roads. Only the proportions of learning mileage spent with passengers were statistically significant (logbook 20% versus control 14%, p=0.04).

4.5.3.2 Hazard perception training intervention

This section presents the results of analysis which seeks to answer the question: for people who engage, what does the hazard perception training intervention change? Hazard perception training engagers and control group participants were compared using the following measures to try to answer this question.

- Self-reported near misses – testing whether there were fewer for those who engaged with the hazard perception training intervention
- Self-reported Driving Events scale score – tested whether this was lower (indicating lower frequency of unsafe driving events) for those who engaged

- Self-reported frequency of breaking the speed limit when driving – tested whether this was less frequent for those who engaged
- Self-reported driving style factors (Q11-22 in test pass and Q35-46 in other surveys)
- Number of theory test attempts before passing – tested whether fewer attempts were needed for those who engaged

All comparisons in this section have been made between 300 participants who engaged with the hazard perception training intervention and the 549 matched control group participants; however, numbers of participants vary by reporting period. See section 4.5.2.2 for the definition of a hazard perception training intervention engager, and a summary of the number of participants in each group included in the analysis of each reporting period.

Near misses

Figure 4-33 and Table 4-22 show the near misses per mile reported by participants who engaged with the hazard perception training intervention and participants in the corresponding control group sample, during each of the reporting periods (months 1-3, months 4-6 and months 7-12). A near miss was described as the impression of only just avoiding an accident.

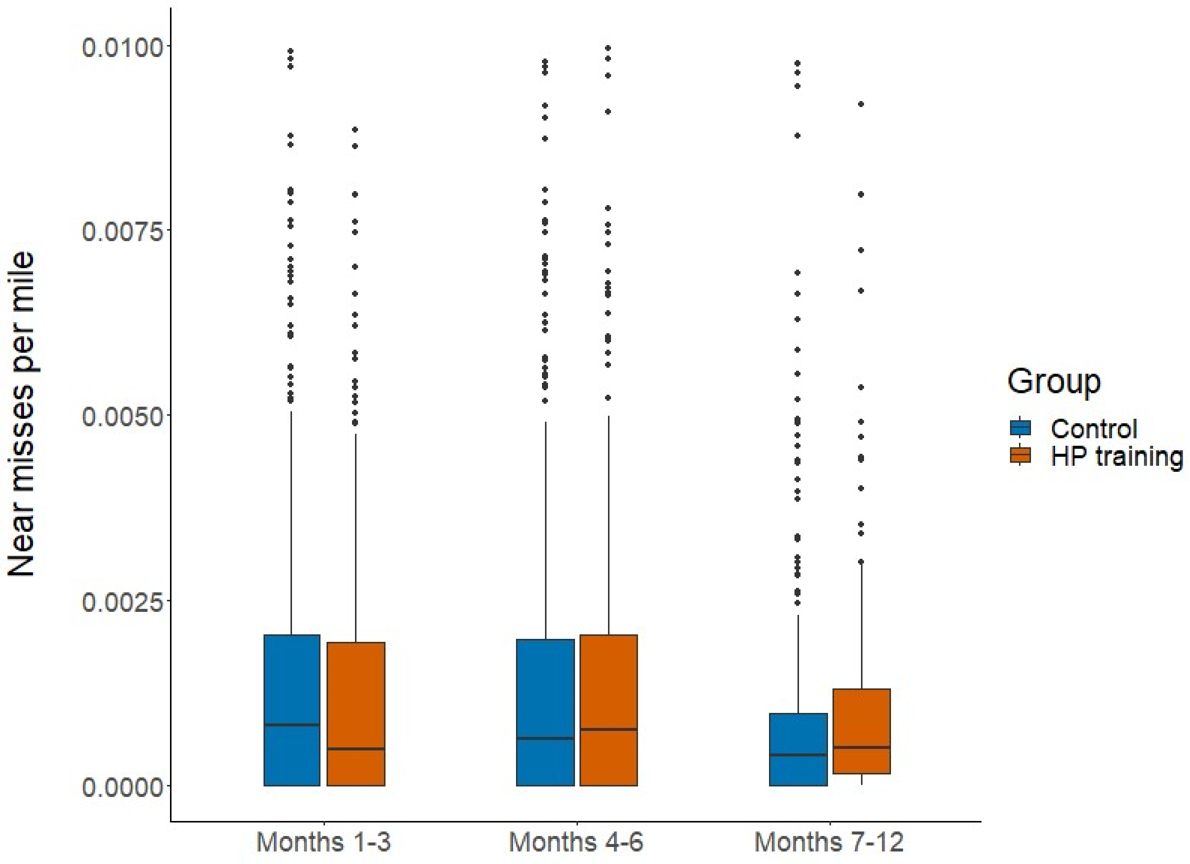


Figure 4-33: Distribution of number of reported near misses per mile in each post-test reporting period for hazard perception (HP) training engagers and the corresponding

control group sample (values >0.01 not shown) (Months 1 to 3: 263 engagers, 445 control group. Months 4 to 6: 253 engagers, 410 control group. Months 7 to 12: 233 engagers, 345 control group).

Table 4-22: Mean number of reported near misses per mile in each post-test reporting period for hazard perception training engagers and the corresponding control group sample (Months 1 to 3: 263 engagers, 445 control group. Months 4 to 6: 253 engagers, 410 control group. Months 7 to 12: 233 engagers, 345 control group).

	Mean number of reported near misses per mile driven		
Group	Months 1-3	Months 4-6	Months 7-12
Control group participants	0.0011	0.0009	0.0005
Hazard perception training engagers	0.0009	0.0010	0.0006

Both control group participants and hazard perception training group engagers reported low numbers of near misses per mile travelled in all reporting periods. The rates per mile are shown in Table 4.20. For context, in months 1-3 and 4-6 people were reporting around one near miss for every 1,000 miles of driving, with this dropping to around one every 2,000 miles of driving in months 7-12. For both groups, the number of near misses per mile decreased over time.

The number of near misses per mile was very similar between groups across all reporting periods. Both groups reported a smaller number of near misses per mile in months 7-12 compared with months 1-3 and months 4-6. Statistical tests showed no significant differences between groups (p=0.57). The difference over time was significant for both groups combined (p<0.01), albeit with a small effect size of 0.01, and the interaction between group and time was not significant (p=0.39), meaning the effect of time was the same for both groups.

For high engagers (compared with the same matched control group sample) only, the statistical results were the same; the differences over time were still significant (p=0.02) with a small effect size of 0.01, but the differences between groups, and the interaction between group and time, were not significant.

Driving events score

In each post-test survey participants were asked about how often each of six driving events deemed unsafe had occurred whilst they were driving during that reporting period. Their responses were converted into the ‘driving events’ score described earlier (section 4.5.3.1), with a lower score indicating a lower frequency of unsafe driving events.

Figure 4-34 shows the distribution of driving events scores for each reporting period for participants who engaged with the hazard perception training intervention and participants from the corresponding control group sample. The corresponding means are shown in Table 4-23.

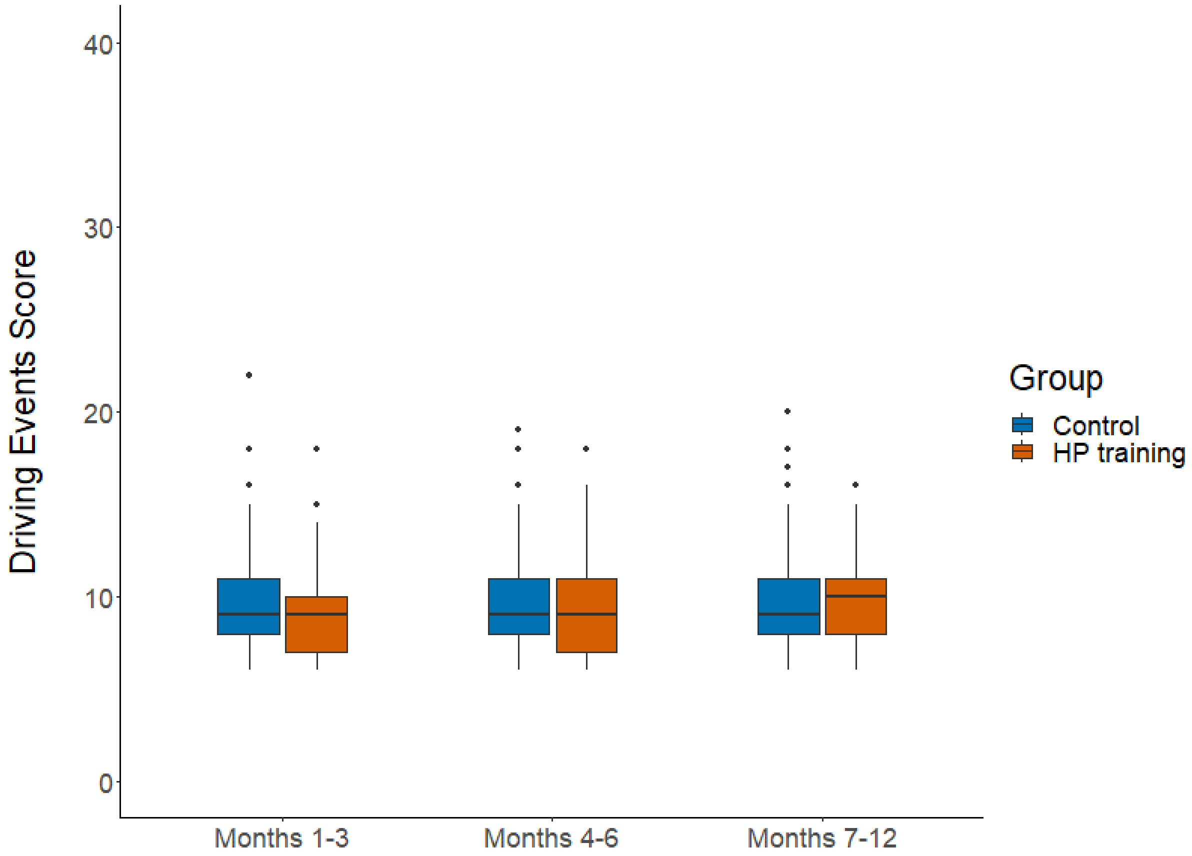


Figure 4-34: Distribution of driving events scores for each post-test reporting period for hazard perception (HP) training engagers and the corresponding control group sample (Months 1 to 3: 263 engagers, 445 control group. Months 4 to 6: 253 engagers, 410 control group. Months 7 to 12: 233 engagers, 345 control group).

Table 4-23: Mean driving event score in each post-test reporting period for hazard perception training engagers and control group participants (Months 1 to 3: 263 engagers, 445 control group. Months 4 to 6: 253 engagers, 410 control group. Months 7 to 12: 233 engagers, 345 control group).

	Mean driving event score		
Group	Months 1-3	Months 4-6	Months 7-12
Control group participants	9.50	9.51	9.75
Hazard perception training engagers	9.14	9.30	9.69

For all reporting periods, the distributions of driving events scores are very similar for participants who engaged with the hazard perception training intervention and participants in the control group sample. This suggests that the hazard perception training intervention has not had an effect on how frequently the events included in this measure occur. For both groups, the mean driving events score increased over time, with lower mean scores in months 1-3 and higher mean scores in months 7-12. This means that, on average, the frequency of the various driving events included in this measure increased over time.

Statistical tests showed no significant difference between groups (p=0.47). The difference over time was significant for both groups (p<0.01), albeit with an effect size of close to zero, and the interaction effect was not significant (p=0.78), indicating that the effect of time was the same for both groups, but this may be a natural effect of, for example, increased confidence leading new drivers to drive more frequently in settings with which they are less familiar.

For high engagers only, the statistical results were the same. The differences over time were still significant for both groups (p<0.01) with a small effect size of 0.01, but the differences between groups, and the interaction between group and time, were not significant, again meaning that the effect of time was the same for both groups.

Exceeding the speed limit

Participants were asked to indicate how often they thought they exceeded the speed limit during each reporting period. Their responses ranged from 0 (never exceeded the speed limit) to 100 (exceeded the speed limit all the time). Figure 4-35 shows the distribution of responses to this question for participants who engaged with the hazard perception training intervention and participants from the corresponding control group sample. The mean frequency scores are shown in Table 4-24.

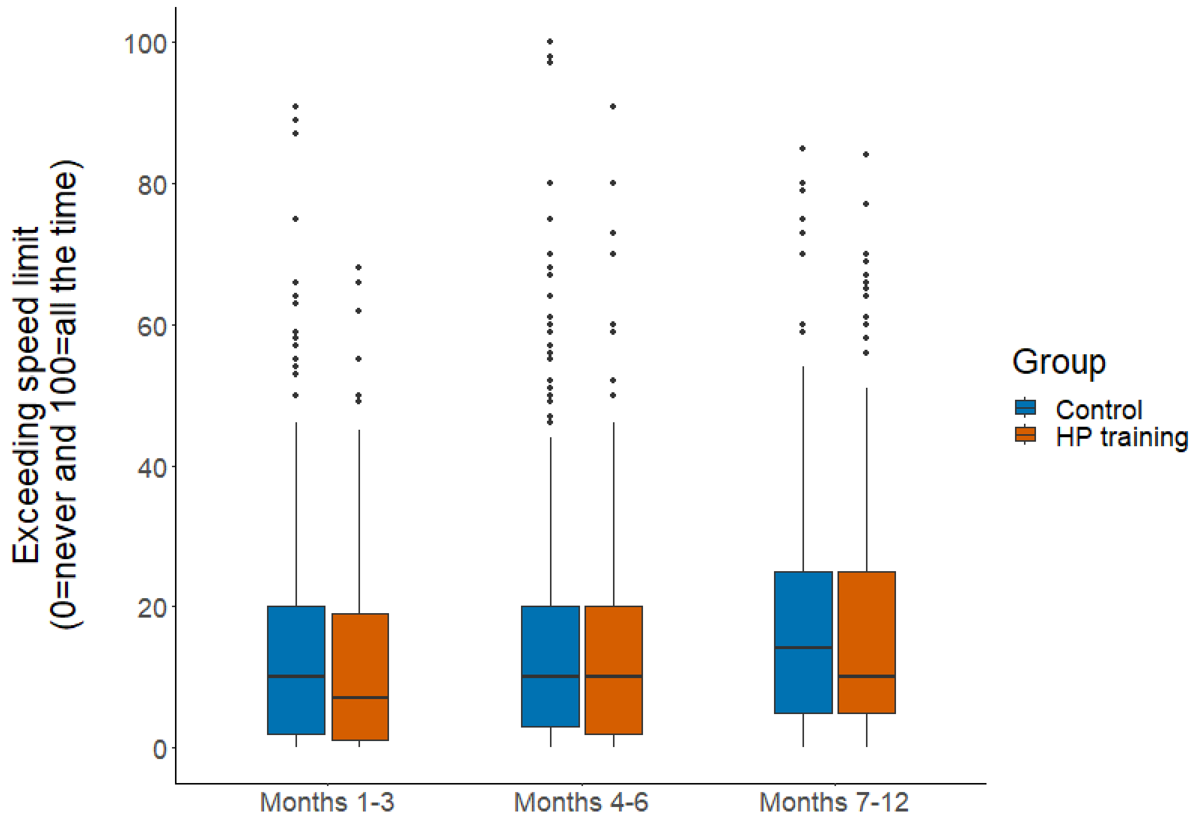


Figure 4-35: Distribution of how often participants thought they exceeded the speed limit for hazard perception (HP) training engagers and the corresponding control group sample (Months 1 to 3: 263 engagers, 445 control group. Months 4 to 6: 253 engagers, 410 control group. Months 7 to 12: 233 engagers, 345 control group).

Table 4-24: Mean exceeding speed limit frequency score in each post-test reporting period for hazard perception training engagers and the corresponding control group sample (Months 1 to 3: 263 engagers, 445 control group. Months 4 to 6: 253 engagers, 410 control group. Months 7 to 12: 233 engagers, 345 control group).

	Mean frequency score		
Group	Months 1-3	Months 4-6	Months 7-12
Control group participants	14.3	15.6	17.4
Hazard perception training engagers	11.4	14.4	17.7

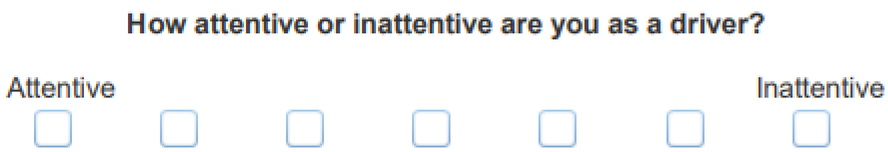
The mean scores for this variable were low in all reporting periods (between 12 and 18 on a scale of 0 to 100). For both groups, the frequency of exceeding speed limits increased slightly over time. There is a noticeably lower score of the hazard perception training group in months 1-3.

The differences in responses between groups was not significant (p=0.75). There was a significant difference in responses over time (p<0.01) with a small effect size of 0.02. The interaction effect was also significant (p=0.045) with a very small effect size of 0.002, indicating that the effect of group was different at each time period (with a difference apparent between the groups at months 1-3 – of around 3 percentage points in the frequency of driving reported as involving exceeding the speed limit). This finding suggests that in very early post-test driving (months 1-3 only) engagement with the hazard perception training intervention was associated with a lower amount of self-reported speeding.

For high engagers only, there were also no statistically significant differences between groups. The effect of time was significant for both groups combined (p<0.01) with a small effect size of 0.01, however the interaction effect was no longer significant (p=0.29), indicating that the effect of time was the same for both groups when the high engagers were considered in isolation.

Driving style

Participants were asked a range of questions about their driving styles shortly after passing their test and in each of the reporting periods (months 1-3, months 4-6, and months 7-12). Driving style was measured using the scales developed by Guppy et al. (1990), which comprise of seven points on 12 bipolar scales such as 'Attentive – Inattentive' or 'Safe-Risky' on which participants were asked to rate themselves. For instance, participants were presented with the driving style depicted below and asked to rate themselves on the scale between the anchors 'attentive' and 'inattentive':



These scales have been used in numerous studies, and the 12 scales typically reduce to three factors characterising particular driving styles. The number of factors is identified using a statistical technique known as factor analysis. This technique is used to simplify data when groups of questions (in this case descriptions of driving style) are thought to be measuring a smaller number of underlying constructs or 'factors'. The logic is that rather than running group comparisons on many items that measure a smaller number of underlying factors, one can establish first what the underlying factors are, and then compare groups on these directly. More information is provided in Appendix C. In this case, a three-factor solution was identified (see Table 4-25). The scale for 'Considerate – Selfish' correlated partially with the first two factors so was removed from the analysis.

Table 4-25: Factor structure for the driving style scales – hazard perception training engagers and matched control group participants

Factor 1	Factor 2	Factor 3
attentive/inattentive	placid/irritable	decisive/indecisive
careful/careless	patient/impatient	experience/inexperienced
responsible/irresponsible	tolerant/intolerant	confident/nervous
safe/risky		fast/slow

As these questions were repeated in each of the surveys, the analysis presents a comparison of driving style over time. To ensure that the results were comparable over time, the coefficients for the factor scores were created using the test pass survey and applied to the responses from each of the reporting periods. This ensured that the results are presented using the same factor coefficients and any differences can be attributed to a change over time. Due to differences in sample sizes over time, the average factor scores are presented for each reporting period.

Figure 4-36 presents the results for Factor 1. A higher score on the scales for Factor 1 indicates a self-reported driving style that is more 'inattentive, careless, irresponsible and risky'.

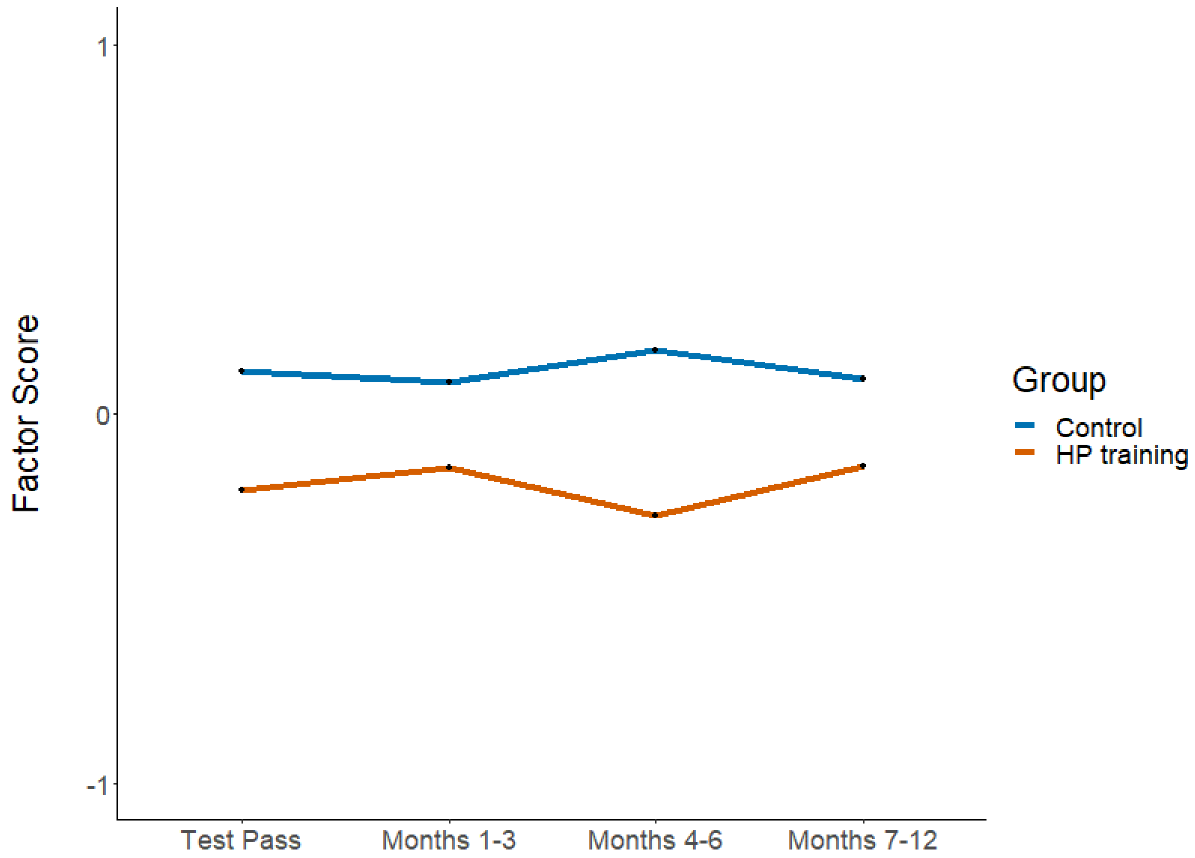


Figure 4-36: Scores for a driving style that is 'inattentive, careless, irresponsible and risky' over time, and by group – hazard perception (HP) training engagers and control group participants (Test Pass: 292 engagers, 525 control group. Months 1 to 3: 263 engagers, 445 control group. Months 4 to 6: 253 engagers, 410 control group. Months 7 to 12: 233 engagers, 345 control group).

Although not a statistically significant effect the hazard perception training group reported a driving style that was slightly less inattentive, careless, irresponsible and risky. Statistical tests showed no significant differences between the groups (*p*=0.13). There were also no differences over time for either group (*p*=0.34) or due to the interaction between group and time (*p*=0.73).

When restricting to only high engagers the results did not change.

Figure 4-37 presents the results for Factor 2. A higher score on the scales for Factor 2 indicates a self-reported driving style that is more 'irritable, impatient and intolerant'.

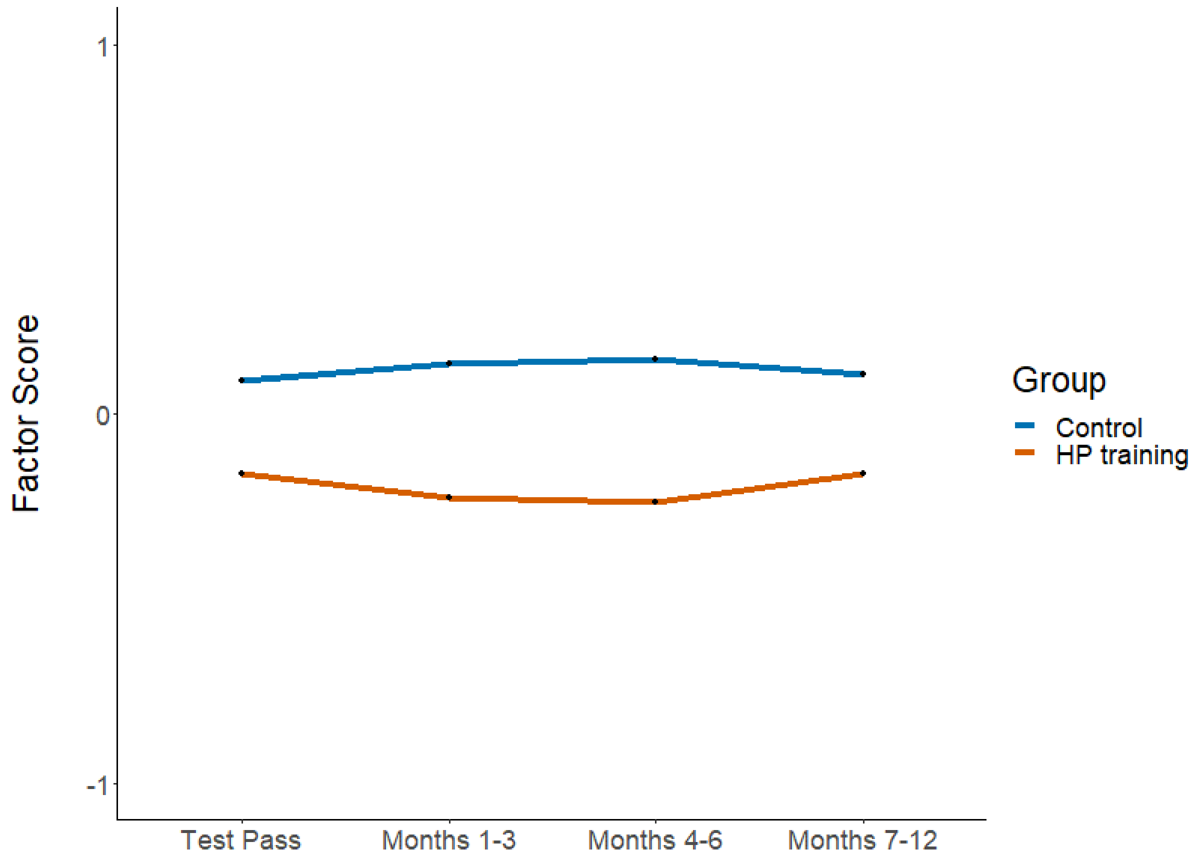


Figure 4-37: Scores for a driving style that is 'irritable, impatient and intolerant' over time, and by group – hazard perception (HP) training engagers and control group participants (Test Pass: 292 engagers, 525 control group. Months 1 to 3: 263 engagers, 445 control group. Months 4 to 6: 253 engagers, 410 control group. Months 7 to 12: 233 engagers, 345 control group).

Although not a statistically significant effect the hazard perception training group reported a driving style that was slightly less irritable, impatient and intolerant. Statistical tests showed no significant difference between the control and hazard perception training groups (p=0.11), or over time for either group (p=0.59). The interaction between group and time was also non-significant (p=0.86).

When restricting to only high engagers, there were no notable changes in the results.

Finally, Figure 4-38 presents the results for Factor 3. A higher score on the scales for Factor 3 indicates a self-reported driving style that is more 'indecisive, inexperienced, nervous and slow'.

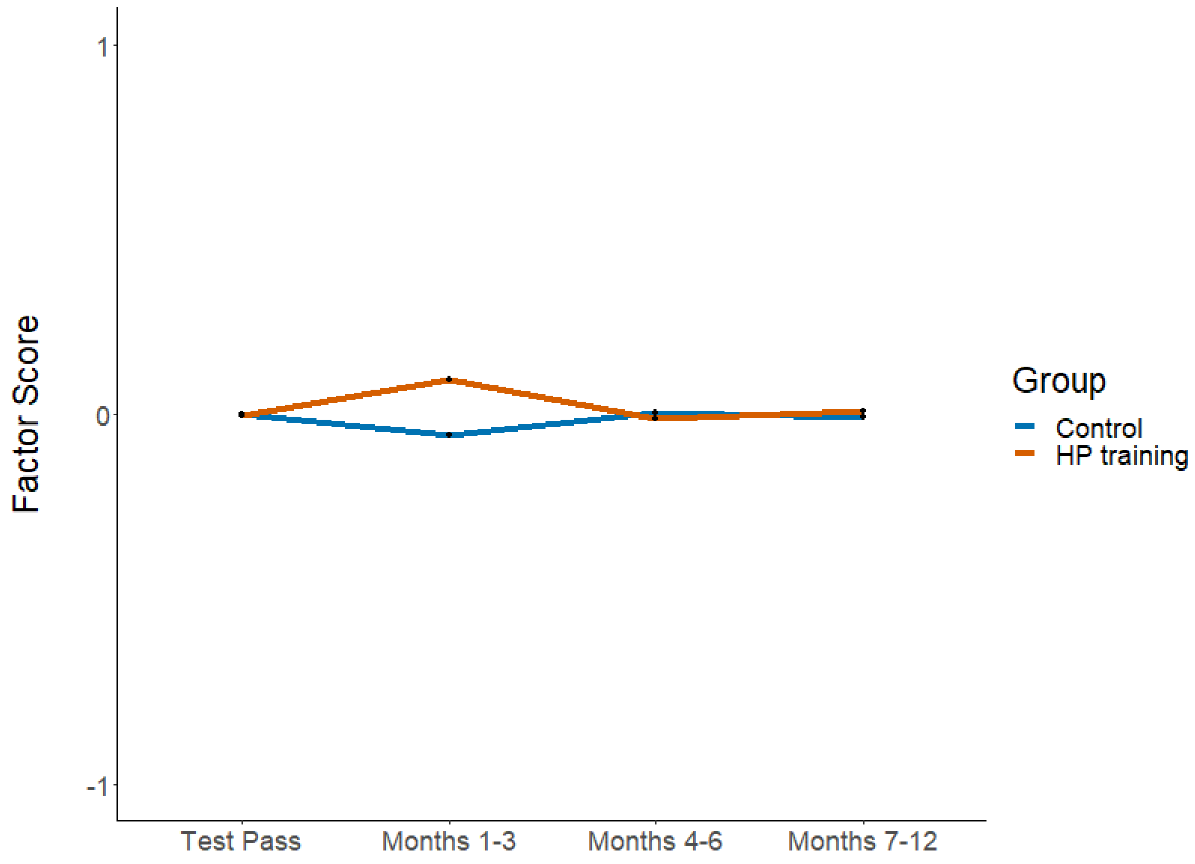


Figure 4-38: Scores for a driving style that is 'indecisive, inexperienced, nervous and slow' over time, and by group – hazard perception (HP) training engagers and control group participants (Test Pass: 292 engagers, 525 control group. Months 1 to 3: 263 engagers, 445 control group. Months 4 to 6: 253 engagers, 410 control group. Months 7 to 12: 233 engagers, 345 control group).

There was no significant difference in self-reported driving style for Factor 3 between groups (p=0.93). The difference over time was not significant for either group (p=0.75) and the interaction effect was not significant (p=0.35), meaning the time effect was the same for each group.

When restricting to high engagers, the results were the same.

Number of theory test attempts

The average number of attempts taken before passing their first theory test, for the hazard perception training engagers and control group participants, is shown in Table 4-26.

Table 4-26 Average number of attempts before passing the theory test – hazard perception training engagers and matched control group participants

Group	Average number of theory test attempts before passing
Control group participants (N = 696)	1.37
Hazard perception training engagers (N = 318)	1.19

There was a statistically significant difference between control group participants and hazard perception training engagers (p<0.01) with a very small effect size of 0.11. This shows that hazard perception training engagers took significantly fewer attempts than control participants.

For high engagers only, there was also a statistically significant difference (p<0.01) between groups. The effect size also increased to 0.17, indicating a small effect. The average number of attempts decreased to 1.15 for high engagers only.

4.5.3.3 Education course intervention

This section presents the results of analysis which seeks to answer the question: for people who engage, what does the education course intervention change? Education course engagers and control group participants were compared using the following measures to try to answer this question.

- Self-reported near misses – testing whether there were fewer for those who engaged with the education course intervention
- Self-reported Driving Events scale score – tested whether this was lower (indicating lower frequency of unsafe driving events) for those who engaged
- Self-reported frequency of breaking the speed limit when driving – tested whether this was less frequent for those who engaged
- Self-reported hours spent driving with an ADI before passing the practical test - tested whether this was more for those who engaged
- Self-reported hours spent driving with another supervising driver (not an ADI) before passing the practical test - tested whether this was more for those who engaged
- Self-reported proportion of learning mileage spent driving on different road types, testing if learning was more spread between road types for education course engagers
- Self-reported driving style factors
- Self-reported proportion of mileage in the dark/with passengers/when tired (Q25/26/32)
- Self-reported attitudes toward speed enforcement and new driver restrictions (Q48)

All comparisons in this section have been made between 82 participants who engaged with the education course intervention and the 440 matched control group participants; however, numbers of participants vary by reporting period. See section 4.5.2.3 for the definition of an education course intervention engager, and a summary of the number of participants in each group included in the analysis of each reporting period.

Near misses

Figure 4-39 and Table 4-27 show the near misses per mile reported by participants who engaged with the education course intervention and participants in the corresponding control group sample, during each of the reporting periods (months 1-3, months 4-6 and months 7-12). A near miss was described as the impression of only just avoiding an accident.

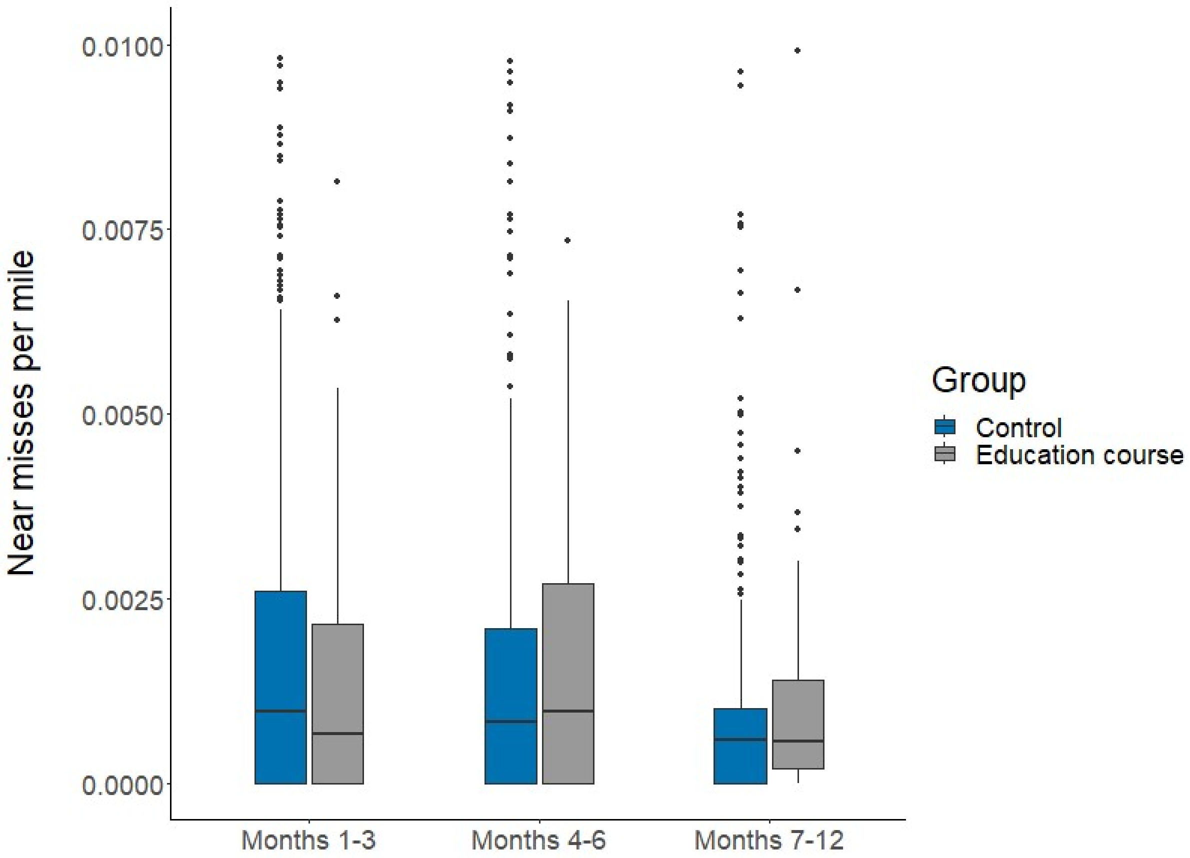


Figure 4-39: Distribution of number of reported near misses per mile in each post-test reporting period for education course engagers and the corresponding control group sample (Months 1 to 3: 71 engagers, 356 control group. Months 4 to 6: 68 engagers, 331 control group. Months 7 to 12: 61 engagers, 277 control group).

Table 4-27: Mean number of reported near misses per mile in each post-test reporting period for education course engagers and the corresponding control group sample (Months 1 to 3: 71 engagers, 356 control group. Months 4 to 6: 68 engagers, 331 control group. Months 7 to 12: 61 engagers, 277 control group).

	Mean number of near misses per mile driven		
Group	Months 1-3	Months 4-6	Months 7-12
Control group participants	0.0012	0.0010	0.0006
Education course engagers	0.0013	0.0009	0.0006

Both control group participants and education course engagers reported low numbers of near misses per mile travelled in all reporting periods. The rates per mile are shown in Table 4.27. For context, in months 1-3 people were reporting around one near miss for every 800 miles of driving, with this dropping to around one every 1,000 miles of driving in months 4-6, and one every 1,600 miles by months 7-12. For both groups, the number of near misses per mile decreased over time. The number of near misses per mile was very similar between groups across all reporting periods.

Statistical tests showed no significant differences between groups (p=0.96). The difference over time was significant for both groups combined (p=0.02), albeit with a small effect size of 0.01, and the interaction between group and time was not significant (p=0.55), meaning the effect of time was the same for both groups.

Driving events score

In each post-test survey participants were asked about how often each of six driving events deemed unsafe had occurred whilst they were driving during that reporting period. Their responses were converted into the 'driving events' score described earlier (section 4.5.3.1), with a lower score indicating a lower frequency of unsafe driving events.

Figure 4-40 shows the distribution of driving events scores for each reporting period for participants who engaged with the education course intervention and participants from the corresponding control group sample. The corresponding means are shown in Table 4-28.

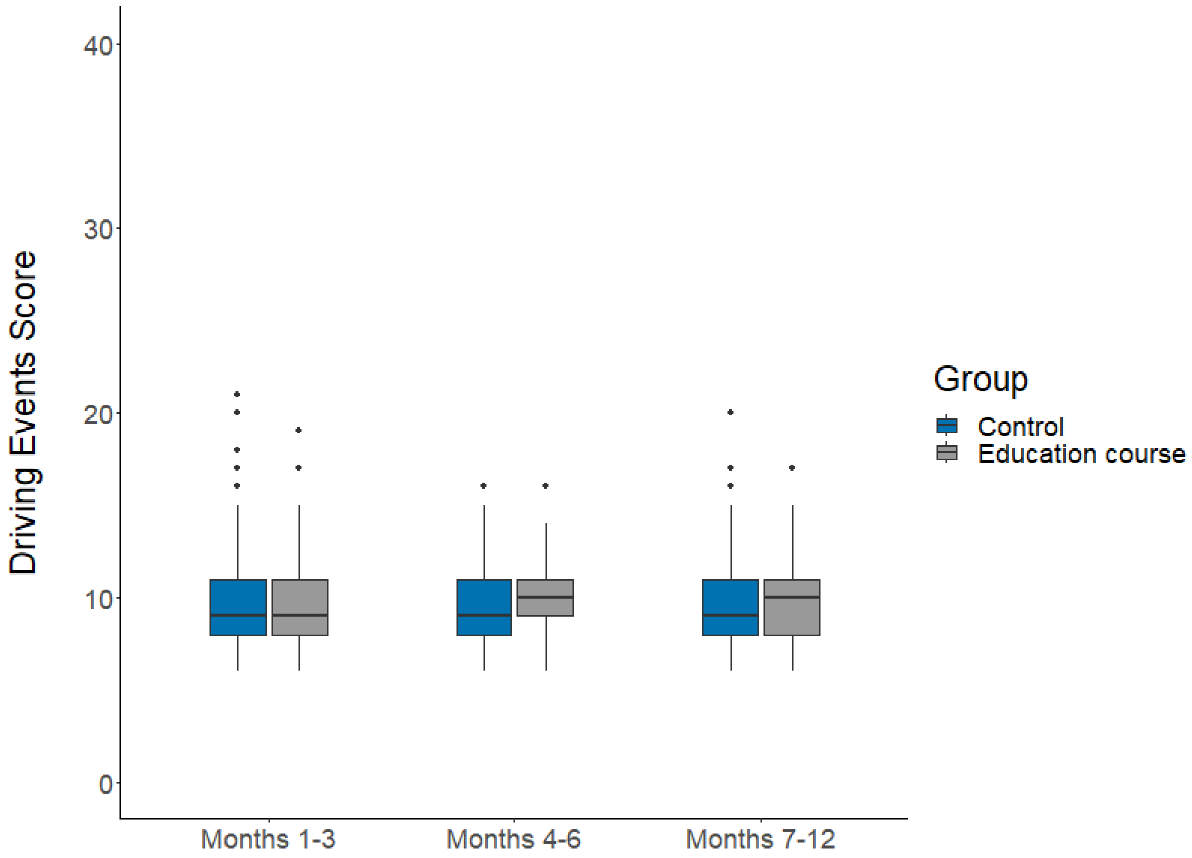


Figure 4-40: Distribution of driving events scores for each post-test reporting period for education course engagers and the corresponding control group sample (Months 1 to 3: 71 engagers, 356 control group. Months 4 to 6: 68 engagers, 331 control group. Months 7 to 12: 61 engagers, 277 control group).

Table 4-28: Mean driving event score in each post-test reporting period for education course engagers and control group participants (Months 1 to 3: 71 engagers, 356 control group. Months 4 to 6: 68 engagers, 331 control group. Months 7 to 12: 61 engagers, 277 control group).

	Mean driving event score		
Group	Months 1-3	Months 4-6	Months 7-12
Control group participants	9.74	9.72	9.95
Education course engagers	9.68	9.96	10.13

For all reporting periods, the distributions of driving events scores are very similar for participants who engaged with the education course intervention and participants in the control group sample. This suggests that the education course has not had an effect on how frequently the events included in this measure occur. For both groups the highest mean scores occurred in months 7-12. This means that, on average, the frequency of the various driving events included in this measure increased over time. Statistical tests showed no significant difference between groups (p=0.58). The difference over time was significant for both groups (p=0.03), albeit with an effect size of close to zero, and the interaction effect was not significant (p=0.73), indicating that the effect of time was the same for both groups.

Exceeding the speed limit

Participants were asked to indicate how often they thought they exceeded the speed limit during each reporting period. Their responses ranged from 0 (never exceeded the speed limit) to 100 (exceeded the speed limit all the time). Figure 4-41 shows the distribution of responses to this question for participants who engaged with the education course intervention and participants from the corresponding control group sample. The mean frequency scores are shown in Table 4-29.

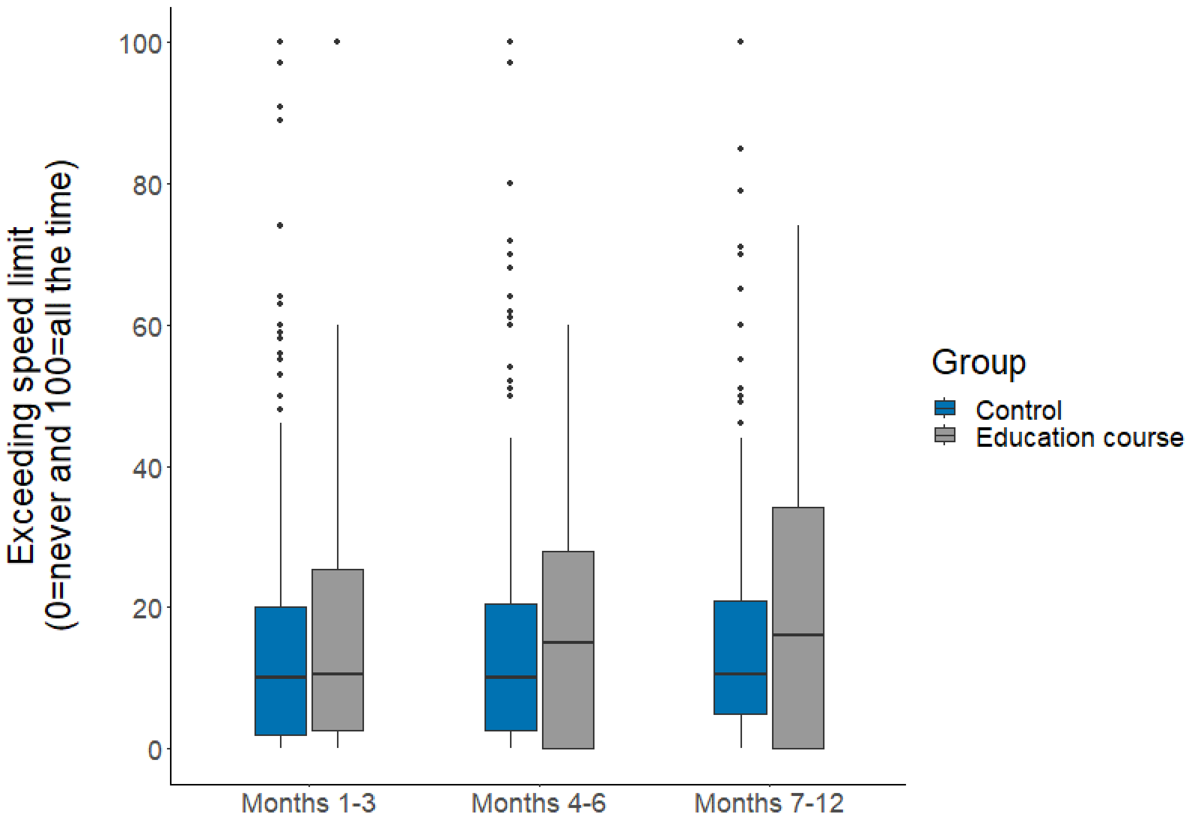


Figure 4-41: Distribution of how often participants thought they exceeded the speed limit for education course engagers and the corresponding control group sample (Months 1 to 3: 71 engagers, 356 control group. Months 4 to 6: 68 engagers, 331 control group. Months 7 to 12: 61 engagers, 277 control group).

Table 4-29: Mean exceeding speed limit frequency score in each post-test reporting period for education course engagers and the corresponding control group sample (Months 1 to 3: 71 engagers, 356 control group. Months 4 to 6: 68 engagers, 331 control group. Months 7 to 12: 61 engagers, 277 control group).

	Mean frequency score		
Group	Months 1-3	Months 4-6	Months 7-12
Control group participants	14.7	15.6	16.6
Education course	15.8	17.3	20.2

For both groups, the frequency of exceeding speed limits increased slightly over time. The differences in responses between groups were not significant (*p*=0.28). There was a significant difference in responses over time for both groups combined (*p*=0.04) with an effect size of close to zero. The interaction effect was not significant (*p*=0.34), indicating that the effect of time was the same for both groups.

Hours spent learning with an ADI and other supervising drivers

In the test pass survey, participants were asked how many hours they had spent driving with an ADI and with another supervising driver before passing the practical test. The mean number of hours with an ADI and another supervising driver for the education course engagers and control group participants is shown in Table 4-30.

Table 4-30: Mean hours reported to be spent learning with an ADI and with another supervising driver for education course engagers and control group participants (79 engagers, 416 control group)

Group	Mean hours with an ADI	Mean hours with another supervising driver	Mean total hours of practice while learning
Control group participants	42.7	29.7	72.4
Education course engagers	41.9	22.9	64.8

Although not statistically significantly different, Table 4-30 shows that the education course engagers spent less time learning to drive with another supervising driver, on average, compared with the control group participants. Education course engagers also spent slightly less time with an instructor. However, these differences were not statistically significant. Statistical tests showed no differences between groups in the number of hours spent with an ADI (p=0.67) or with another supervising driver (p=0.21).

Proportion of mileage learning in different conditions

In the test pass survey, participants were asked about the number of hours spent driving in different conditions before passing the practical test, including on different road types, with additional passengers (not including the supervising driver), in the dark, and in wet conditions. The proportion spent learning in each condition has been averaged across the participants in each group.

Figure 4-42 shows the proportion of learning miles reported to be spent driving in different conditions for education course engagers and control group participants.

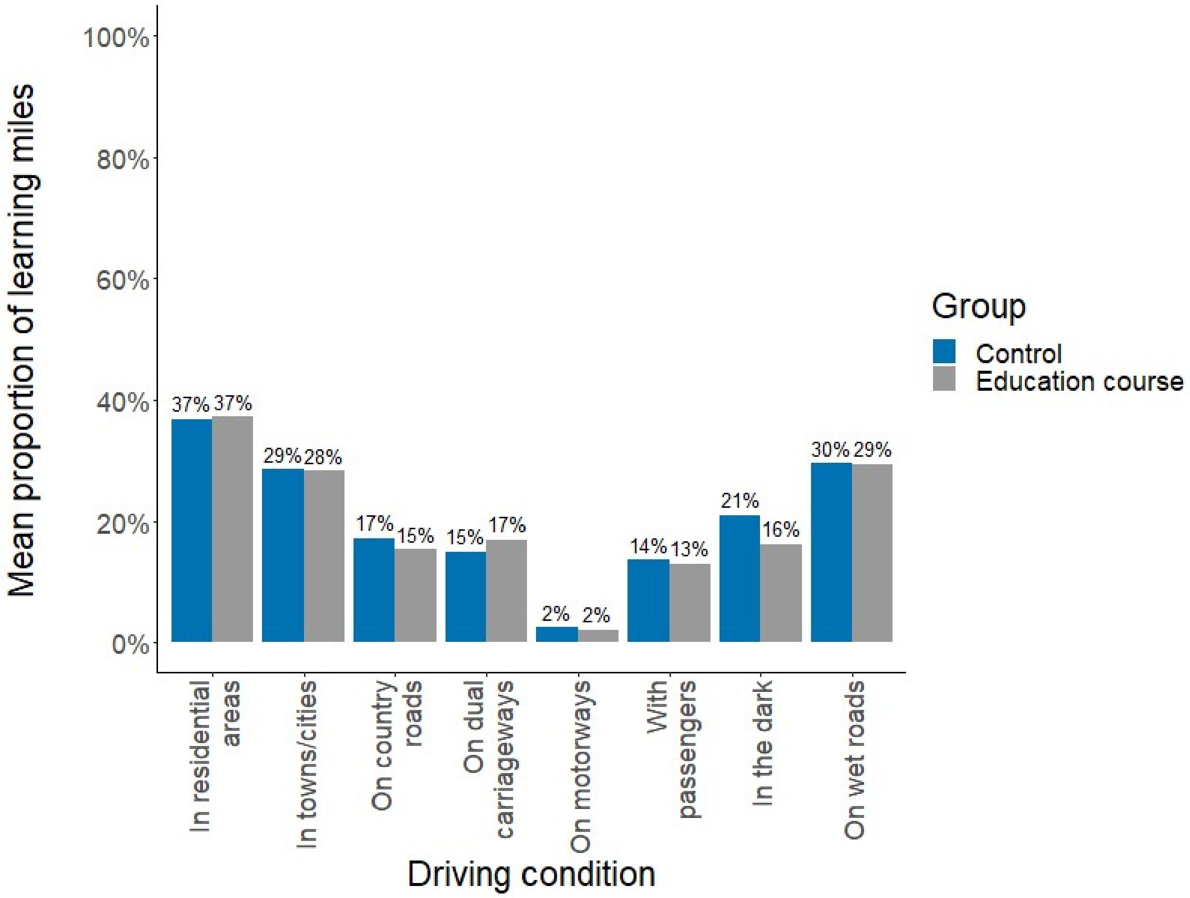


Figure 4-42: Proportion of total learning miles reported to be spent driving in different conditions for education course engagers and the corresponding control group sample (79 engagers, 416 control group).

The proportion of miles spent learning is similar across most of the different conditions in Figure 4-42 between education course engagers and the control group. The most notable difference is in the proportion of miles spent learning in the dark – this proportion was significantly (p=0.03) higher for the control group participants (21% versus 16%). The difference for dual carriageways was also significant (p=0.04) with those on the education course reporting a higher mean proportion of their learning miles as being on dual carriageways than the control group (17% versus 15%).

Driving style

As in section 4.5.3.2 (which describes the driving style questions in more detail), the twelve driving style question scales were suitable for factor analysis and a three-factor solution was identified (see Table 4-31). The scale for ‘Considerate – Selfish’ correlated partially with the first two factors.

Table 4-31: Factor structure for the driving style scales – education course engagers and matched control group participants

Factor 1	Factor 2	Factor 3
attentive/inattentive	placid/irritable	decisive/indecisive
careful/careless	patient/impatient	experience/inexperienced
responsible/irresponsible	tolerant/intolerant	confident/nervous
safe/risky		fast/slow

To ensure that the results were comparable over time, the coefficients for the factor scores were created using the test pass survey and applied to the responses from each of the reporting periods. This ensured that the results are presented using the same factor coefficients and any differences can be attributed to a change over time. Due to differences in sample sizes over time, the average factor scores were used for each reporting period.

In no cases were there statistically significant differences between the control participants and the education engagers, or changes over time.

Driving with similar aged passengers and in the dark

Participants were asked what proportion of their mileage in each reporting period had been driven on various road types and in various other conditions. There were two driving conditions which were particularly relevant to the education course intervention: driving with at least one passenger of a similar age and driving in the dark. The proportion of mileage driven in each of these conditions in each reporting period is compared between education course engagers and control group participants in Figure 4-43.

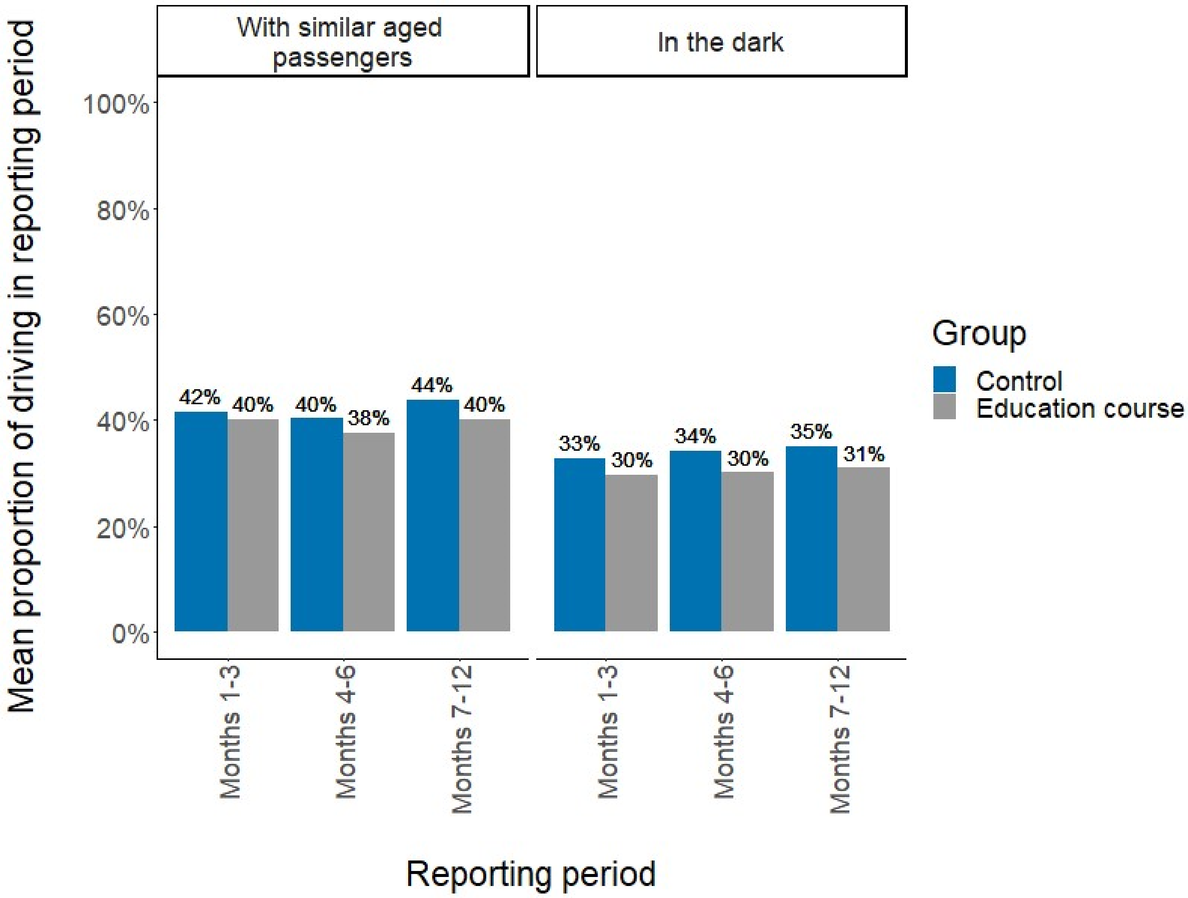


Figure 4-43: Proportion of miles driven by education course engagers and control group participants reported to be with passengers and in the dark in each post-test reporting period (Months 1 to 3: 71 engagers, 356 control group. Months 4 to 6: 68 engagers, 331 control group. Months 7 to 12: 61 engagers, 277 control group).

Although there were some slight differences between the groups for these measures, none of the pairwise comparisons were statistically significant.

Driving when tired

Participants were asked to indicate how often they thought they drove when tired during each reporting period. Their responses ranged from 0 (never driving when tired) to 100 (driving when tired all the time). Figure 4-44 shows the distribution of responses to this question for participants who engaged with the education course intervention and participants from the corresponding control group sample. The mean frequency scores are shown in Table 4-32.

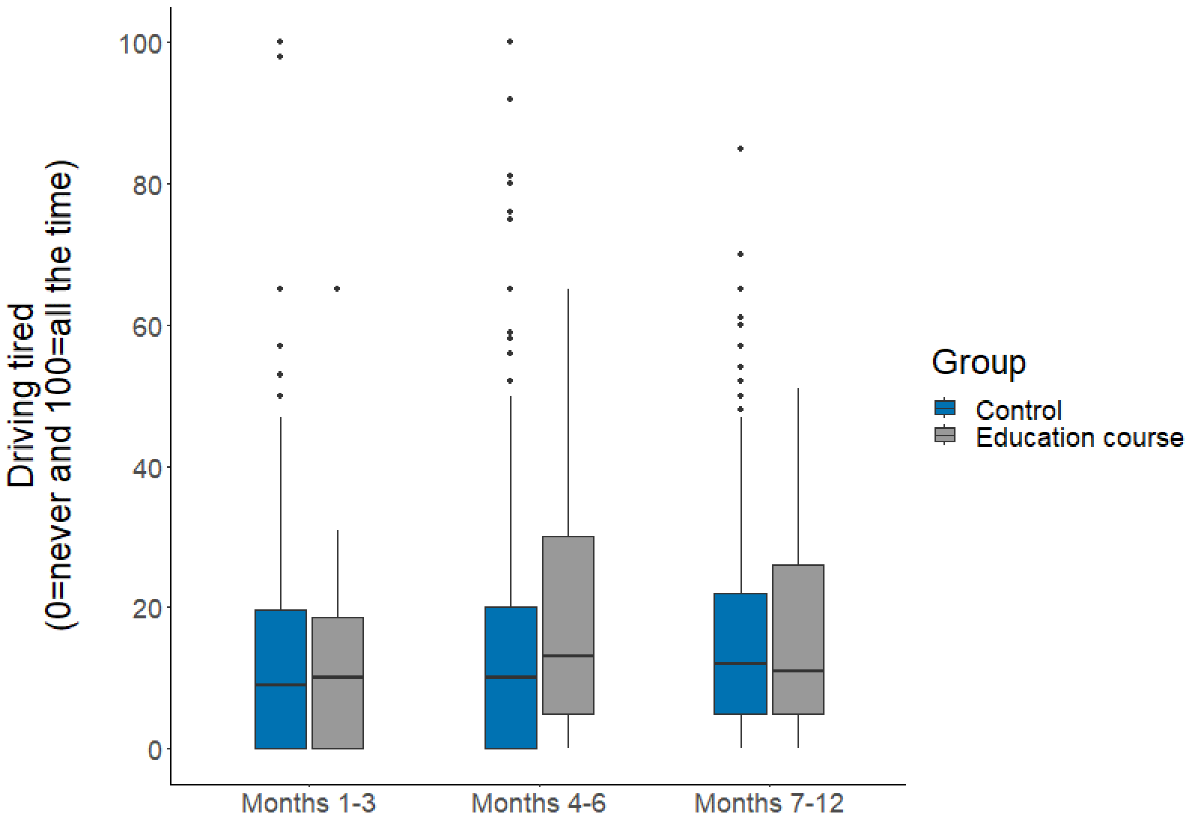


Figure 4-44: Distribution of how often participants thought they drove when tired for education course engagers and the corresponding control group sample (Months 1 to 3: 71 engagers, 356 control group. Months 4 to 6: 68 engagers, 331 control group. Months 7 to 12: 61 engagers, 277 control group).

Table 4-32: Mean driving when tired frequency score in each post-test reporting period for education course engagers and the corresponding control group sample (Months 1 to 3: 71 engagers, 356 control group. Months 4 to 6: 68 engagers, 331 control group. Months 7 to 12: 61 engagers, 277 control group).

	Mean frequency score		
Group	Months 1-3	Months 4-6	Months 7-12
Control group participants	12.6	14.3	16.2
Education course	11.2	18.3	15.6

The mean scores for this variable were low in all reporting periods (between 11 and 19 on a scale of 0 to 100). For both groups, the frequency of driving when tired increased over time. Although not statistically significantly different, there is a noticeably higher score of the education course group in months 4-6.

The differences in responses between groups was not significant (p=0.28). There was a significant difference in responses over time (p<0.01) with a small effect size of 0.01. The interaction effect was not significant (p=0.28), indicating that the effect of time was the same for both groups.

Attitudes towards new driver restrictions and more general enforcement

After passing their driving tests and in each of the reporting periods (months 1-3, months 4-6 and months 7-12) participants were asked to select their level of agreement towards speed enforcement and new driver restrictions. These included 7-point Likert-scale statements around stronger enforcement of speed limits, drink-driving and alcohol, minimum practice hours and more supervision for six months after passing their test.

The data were suitable for factor analysis (as described in section 4.5.3.2 and Appendix C) and a three-factor solution was identified as shown in Table 4-33. The question 'For 6 months after passing their test, new drivers should not be allowed to drink any alcohol if driving' did not correlate strongly with any of the three factors. This item was, therefore, not used.

As these questions were repeated in each of the surveys, the analysis presents a comparison of how driver attitudes towards speed enforcement and new driver restrictions change over time. To ensure that the results were comparable over time, the coefficients for the factor scores were created using the test pass survey and applied to the responses from each of the reporting periods. This ensured that the results are presented using the same factor coefficients and any differences can be attributed to a change over time.

For each of the three factors, the education course engagers scored slightly higher than the matched control group across all time points (indicating slightly more agreement with the suggested enforcement and restrictions). However, there were no statistically significant differences between the control and education course groups. There were also no statistically significant changes over time.

Table 4-33: Factors for the attitudes to driver restrictions and enforcement scale

Factor 1 – attitudes to learner driver requirements	Factor 2 – attitudes to general enforcement	Factor 3 – attitudes to post-test restrictions
Learner drivers should have a minimum learning period of at least 12 months before they can take their test Learner drivers should need to have at least 100 hours of on-road practice before they can take their test	Speed limits should be more strongly enforced Drink/drug driving laws should be more strongly enforced Laws to stop the use of mobile phones and other distracting devices when driving should be more strongly enforced	For 6 months after passing their test, new drivers should not be allowed to drive in the dark without a supervising driver For 6 months after passing their test, new drivers should not be allowed to carry passengers under 25 without a supervising driver

5 Discussion and conclusions

The learner driver arm of the Driver2020 project aimed to test the real-world effectiveness of three interventions, on both collision involvement, and a number of surrogate safety measures, in the first 12 months of post-test driving.

The main findings from the study are discussed below, first for the effect of the interventions on collisions, and then on surrogate measures. It should be noted that the COVID-19 pandemic affected both learning to drive, and post-test driving during the study for the majority of participants. Thus, some caution should be exercised in how generalisable the findings are to future circumstances.

5.1 Effect of interventions on collisions

Engagement with the three interventions was low, ranging from 3% to 11% of participants who were offered the intervention actually engaging to some degree.

The intention to treat analysis, which compared collision involvement rates between the intervention and control groups for all participants (regardless of engagement) showed that the groups did not differ in collision risk after mileage, age, gender, and COVID-19 effects were controlled. This is not surprising given the low engagement rates, since even if the interventions were extremely effective for those who engaged, they would only be having an effect on a small proportion of the intervention group. The analogy is that if a medicine exists that cures a disease with 100% efficacy, and only 1% of people take the medicine, then the real-world effectiveness of the medicine will be 1%. Unfortunately, it is not possible to examine differences in collisions solely in the people who engaged with the interventions, as the samples sizes were insufficient to support such an analysis; this is why surrogate measures were used for the per-protocol analysis (see below).

Based on the trial sample, this finding suggests that if the logbook, hazard perception training and education interventions trialled the Driver2020 study were offered on a similar voluntary basis to learner drivers aged 17-24 in Great Britain they would be unlikely to lead to any measurable reduction in collision risk in this population.

5.2 Effect of interventions on surrogate measures

To examine the potential impacts on those participants who did engage, groups were compared on a number of surrogate measures of risk. This ‘per protocol’ analysis included only those who engaged with the three interventions, and corresponding control participants. Due to the sample sizes being smaller than anticipated in the engager groups, extra caution needs to be exercised in interpretation of non-significant effects. It is possible that some differences that did not reach statistical significance would have done so had the sample sizes been more in line with what was expected (through more participants having the opportunity to pass their test and provide data, or through greater engagement). Such examples are discussed below.

The amount of engagement with interventions in those who did engage was also relatively low, with the ‘threshold’ for being described as an engager typically ‘any use’ of the intervention in question. This needs to be considered when interpreting the results; because engagement ‘dose’ was low even in those people who did engage, it may be that the effects seen were limited by this.

5.2.1 Logbook intervention

Engagement with the logbook intervention was associated with a statistically significant change in one surrogate measure. This was the proportion of learning reported as being done with passengers in the car. Logbook engagers, when compared with control participants, reported a greater proportion of their time practising with passengers. Despite the statistical significance, the effect size was very small. It is also not clear what impact this would have on safety, although potentially it could have a safety benefit through increasing the degree of overlap between the training (learning to drive) and transfer (post-test driving) contexts (Groeger & Banks, 2007; Barnett & Ceci, 2002). It is also potentially a spurious effect; the logbook group as a whole (even non-engagers) had a slightly higher rate of *post-test* driving with passengers than the control group, which may mean that they are more likely to have passengers in the car for reasons not associated with the intervention.

While not statistically significant, there was also a trend in the data for logbook engagers to have a higher total number of hours of practice relative to the control group (76.3 hours in the control group versus 81.8 hours in the logbook engagers – combined practice with ADI and other supervising driver). This difference is encouraging, as increased on-road practice during the learning stage would be expected to have a safety benefit, but the difference did not reach statistical significance, potentially due to the lack of statistical power caused by the lower-than-intended participant numbers. It is worth considering the practical significance of such a difference in this case. Regardless of whether this increase might have reached statistical significance had the sample been larger, the fact remains that this increase is still not the target of 100 hours set within the study for this group; this target was based on evidence reviewed in previous work (see work reviewed in Pressley et al., 2016) suggesting that a minimum of 100 hours of pre-test practice is likely to be needed before any safety benefits are seen post-test.

The logbook intervention findings suggest that such an app can lead to very modest changes in the learning to drive of those who engage (in the proportion of learning with passengers).

5.2.2 Hazard perception training intervention

Engagement with the hazard perception training intervention was associated with statistically significant changes in two surrogate measures. First, engagers (although not the subset of ‘high engagers’ – see below) showed a reduction in the frequency with which they reported driving above the speed limit, in their first three months of post-test driving (from 14.3 to 11.4 where ‘0’ means ‘not at all’ and ‘100’ means ‘all the time’). This is a very encouraging result for safety given the high correlation between driving at higher speeds and collision risk (Elvik, Vadeby, Hels & Van Schagen, 2019). It is also aligned with previous findings in the hazard perception literature. For example, both McKenna, Horswill and Alexander (2006) and Helman, Palmer, Delmonte and Buttress (2012) showed that drivers (and riders) with higher levels of hazard perception skill chose lower speeds in response to hazardous road situations. The fact that this effect did not persist when only ‘high engagers’ in the hazard perception training group were included in the analysis is slightly puzzling, although again may have to do with the low sample sizes involved – less than half of the 'engagers' sample in this case.

The second significant change was that engagers with the hazard perception training intervention took fewer attempts on average to pass their driving theory test (of which the hazard perception test is one part, along with the multiple-choice component). This finding would be predicted if the hazard perception training intervention increased engagers' hazard perception skill, as it would lead to them scoring higher on the test; it was also maintained in the high engager group. Having a higher level of hazard perception skill is desirable for safety, with many studies demonstrating this (for example Wells et al., 2008).

While not reaching statistical significance, two of the driving style factors ('inattentive, careless, irresponsible and risky' and 'irritable, impatient and intolerant') showed a trend for a relationship with hazard perception training engagement (both in a safer direction). In the context of the lower sample sizes in the learner arm of the study (and therefore the lower than anticipated statistical power) there is the potential for identifying a relationship between hazard perception training and some driving styles with a larger sample size.

The hazard perception training intervention findings suggest that such an intervention delivered during learning to drive could have safety benefits if a way were found to ensure that people engage with it. Hazard perception is a skill known to be amenable to training and to be associated with collision risk; see Cao et al. (2022) for a recent review and Grayson and Sexton (2022) for a review of early UK work leading to the development of the hazard perception test.

5.2.3 Education intervention

Engagement with the education intervention was associated with statistically significant changes in two surrogate measures. The first was that engagers reported a lower proportion of their learning in the dark. The second was an increase in the proportion of their learning reported as being on dual carriageways. The effect sizes for these differences were small. The safety impact of these is unclear; a wider range of practice contexts during the learning stage is generally accepted as being a good thing for later transfer of experience, not only in driving (Groeger & Banks, 2007) but in all skill learning (Barnett & Ceci, 2002), so the argument could be made that changes in learning may have some safety impacts later, although more research would be needed to identify boundary conditions for this.

The very low engagement with this intervention also means that the sample was much smaller than anticipated. Given this context it is also worth considering two observed differences that failed to reach statistical significance but where the *p*-value was close to the significance threshold. The first was that education engagers reported less time learning with a supervising driver. The second was that they reported a higher frequency of their driving post-test as involving speeding. Both of these trends, and especially the latter, if identified in a larger sample as being statistically significant, would have been undesirable for safety.

5.3 Other findings of note

5.3.1 Impact of COVID-19

COVID-19 had a major impact on the Driver2020 study, in particular in terms of the participant numbers in the learner arm, and also seems to have had an impact on the risks faced by newly qualified drivers when they begin driving. In the learner arm of the study, those who passed their driving test before the pandemic began in March 2020 were up to 5.1 times more likely (depending on mileage) to report a collision in their first year of post-test driving than those who passed after this point.

There are two things about this finding that bear further examination, and future work is planned to look at this issue in more detail. First, that there was any difference at all between those passing pre- and post-pandemic is itself interesting. Second, there was a very large increase in the effect in the learner arm (this report) compared with the novice arm (Weekley et al., 2024a) for some participants (depending on mileage).

Regarding the existence of the effect, there are likely several mechanisms at play. One is that people who passed their test after the pandemic by definition were more likely to experience a delay in licensure, and thus be both older and potentially have more on-road practice when they finally began driving; as noted in Section 1 and Appendix A.2 both of these changes would be expected to lead to lower collision risk, as observed (Wells et al., 2008; Forsyth et al., 1995; Maycock et al., 1991). Another potential contributing factor is that the driving environment itself was fundamentally different immediately after the pandemic began, with national lockdowns leading to changes in traffic characteristics. For example, in many countries (including Great Britain) average traffic speeds increased during the initial lockdowns (Wegman & Katrakazas, 2021). While the changes would be expected to increase risk for all drivers, newly qualified drivers might be especially prone to the increased task demand that higher speeds would bring (Fuller, 2011). Further analyses of the data would be required to understand the detailed impact of the COVID-19 pandemic, and the mechanisms by which it occurred, on newly qualified drivers in Great Britain during this period.

The difference in risk between the novice and learner arms of the study is more puzzling. In the novice arm of the study those who passed their test before the pandemic were 1.26 times more likely to report a collision in their first year of driving than those who passed after (Weekley et al., 2024a). In the learner arm (this report) the increase was up to 5.1 times. The interaction with mileage in the learner arm is another topic that would require further investigation to understand as this interaction effect was not present in the novice arm. This effect was such that each additional mile of driving added less risk for those passing before the pandemic than for those passing after. Again, further exploration of the data would be needed to understand the mechanisms involved.

5.3.2 The effects of post-test experience

In both the novice arm (Weekley et al., 2024a) and learner arm (this report) of the study, the analysis showed that as mileage increased, although the number of self-reported collisions went up, the *rate* of increase per mile went down. This confirms the importance of post-test experience in reducing risk,, building on the evidence base summarised in Section 1 and Appendix A.2. One finding of note however in the learner arm was that the coefficient was larger than in the novice arm (-0.63 compared with -0.40). This suggests that for people in the learner arm of the study, each additional mile of experience bought greater safety gains than it did for the people in the novice arm of the study.

It is possible this is related to the fact that novices were recruited when they passed their test and from this point were guaranteed to have the opportunity to provide survey data, while learners (having been recruited when they began learning) needed to have the opportunity to pass their test and then provide survey data, something that was not guaranteed due to the long delays in driving test access after the pandemic. While the delays for tests were widespread novices recruited into the study were by definition those who had been *able* to book and complete tests, while learners were subject to the chances of further delays once they were ready to book, and also ran the risk of needing to gain refresher training with the interruption of their learning period. In short, the novices and learners in the study may have been people with different ‘types’ of experience, although in the current report how these ‘types’ of experience might be defined has not been explored.

This difference in recruitment route may also have some bearing on the differences in the risk-increasing effect of COVID-19 between the arms; it is possible for example that learners passing after the pandemic in the Driver2020 study experienced greater licensure delay than novices recruited, which may have amplified any safety benefits associated with being older and more experienced when they began driving. How this might have impacted on the differences in the beneficial effects of experience between the two groups would also require further investigation.

5.4 Conclusion – learner driver arm

None of the three interventions offered to learner drivers reduced collisions in the first 12 months of post-test driving when offered under a voluntary approach.

An important finding that helps in the interpretation of this main conclusion is that when such interventions are offered under the conditions of voluntary engagement used in this study, levels of engagement are extremely low.

Analysis of surrogate measures with the people who engaged to some degree suggests that small to modest changes in some variables are possible, particularly from the hazard perception training intervention.

None of the interventions show any sign of improving safety for young and novice drivers to the extent shown by stronger, legislative approaches such as the introduction of hazard perception testing in Great Britain in 2002 (Wells et al., 2008), or stronger approaches to licensing seen in other countries (Russell et al., 2011; Kinnear et al., 2013).

Further work with the Driver2020 dataset is expected to help further elucidate the effects of the COVID-19 pandemic on learning to drive and on early post-test driving in young and novice drivers. It also provides a recent dataset that can be used to examine this group, and the long-understood road safety challenge they present.

References

Barnett SM and Ceci SJ (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612-637.

Begg D and Stephenson S (2003). Graduated driver licensing: the New Zealand experience. *Journal of Safety Research*, 34(1), 99-105.

Brown ID, Groeger JA and Biehl B (1987). Is driver training contributing enough towards road safety? In J. A. Rothengatter & R. A. de Bruin (Eds.), *Road Users and Traffic Safety*. Assen: Van Gorcum.

Cao S, Samuel S, Murzello Y, Ding W, Zhang X & Niu J (2022). Hazard perception in driving: a systematic literature review. *Transportation Research Record*, 2676(12), 666-690.

Christie R (2001). *The effectiveness of driver training as a road safety measure. A review of the literature.* Victoria, Australia: Royal Automobile Club of Victoria.

Clarke DD, Ward P and Truman W (2002). *In-depth accident causation study of young drivers* (TRL542). Crowthorne: Transport Research Laboratory.

Clinton K and Lonero L (2006). *Guidelines for evaluating driver education programs.* Washington, DC: AAA Foundation for Traffic Safety.

Department for Transport (2015). *Working together to build a safer road system: British road safety statement.* Stationary Office. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/487949/british_road_safety_statement_web.pdf

Department for Transport (2022). https://www.gov.uk/government/statistical-data-sets/driving-test-statistics-drt, accessed 12/01/22

Elvik R, Vadeby A, Hels T and Van Schagen, I. (2019). Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. *Accident Analysis and Prevention*, 123, 114-122.

Forsyth E, Maycock G and Sexton B (1995). *Cohort study of learner and novice drivers. Part 3: Accidents, offences and driving experience in the first three years of driving.* Crowthorne: Transport Research Laboratory.

Fuller R (2011). *Driver control theory: From task difficulty homeostasis to risk allostasis*. In Handbook of traffic psychology (pp. 13-26). Academic Press.

Fylan F (2017). *Using Behaviour Change Techniques: Guidance for the road safety community.* RAC Foundation.

Glad A (1988). *Phase 2 driver education, effect on accident risk.* Oslo, Norway: Transport Institute.

Goldstein LG (1972). Youthful drivers as a special safety problem. *Accident Analysis and Prevention*, *4*(3), 153-189.

Gov.uk website deprivation data (2019). [online]. [Accessed March 2022]. Available from World Wide Web: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/467774/Fi le_7_ID_2019_All_ranks__deciles_and_scores_for_the_Indices_of_Deprivation__and_population_denominators.csv/preview

Gov.scot website deprivation data (2020). [online]. [Accessed March 2022]. Available from World Wide Web: https://www.gov.scot/publications/scottish-index-of-multiple-deprivation-2020v2-data-zone-look-up/

Gov.wales website deprivation data (2019). [online]. [Accessed March 2022]. Available from World Wide Web: https://statswales.gov.wales/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation

Grayson GB and Sexton BF (2002). *The development of hazard perception testing* (TRL558). Crowthorne: Transport Research Laboratory.

Groeger, JA and Banks AP (2007). Anticipating the content and circumstances of skill transfer: Unrealistic expectations of driver training and graduated licensing?. Ergonomics, 50(8), 1250-1263.

Guppy A, Wilson P and Perry J (1990). Driving attitudes and driving experience. In Benjamin, T. (ed.) Driving Behaviour in a Social Context. Caen, France: Paradigme

Helman S, Grayson GB and Parkes AM (2010). *How can we produce safer new drivers? A review of the effects of experience, training, and limiting exposure on the collision risk of new drivers* (INS005). Crowthorne: Transport Research Laboratory.

Helman S, Palmer M, Delmonte E and Buttress S (2012). *Development of a video measure of hazard perception skill and a group discussion-based hazard perception training package for motorcyclists* (PPR615). Crowthorne: Transport Research Laboratory.

Helman S, Wallbank C, Chowdhury S, Hammond J, Kinnear N, Buttress S, Jenkins R and Grayson G (2017). *Transforming the practical driving test: final report* (PPR828). Crowthorne: Transport Research Laboratory.

Helman S and Weekley J (2024). *Driver2020 – an evaluation of interventions designed to improve safety in the first year of driving. Report D4: Summary of findings* (PPR2012). Commissioned and funded by the Department for Transport. Crowthorne: Transport Research Laboratory.

Hitchings J, Holcombe A, Christie N, Weekley J and Helman S (2024). *Driver2020 - an evaluation of interventions designed to improve safety in the first year of driving. Report D3: Delivery of interventions and engagement by novice and learner drivers* (PPR2011). Commissioned and funded by the Department for Transport. Crowthorne: Transport Research Laboratory.

House of Commons (2021). Road safety: young and novice drivers. Fourth Report of House of Commons Transport Committee Session 2019-2021, HC169. London: House of Commons Transport Committee.

Isler RB and Starkey NJ (2008). The ‘frontal lobe’ project: A double-blind, randomized controlled study of the effectiveness of higher level driving skills training to improve frontal lobe (executive) function related driving performance in young drivers. Final report by the Traffic and Road Safety Research Group, Psychology Department, University of Waikato. Hamilton, New Zealand: University of Waikato.

Jones B (1993). *The effectiveness of skid-car training for teenage novice drivers in Oregon.* Salem, OR: Drive and Motor Vehicle Services.

Katila A, Keskinen E, Hatakka M and Laapotti S (1996). Skid training for novice drivers – benefit for adults, pitfall for youngsters. *Paper presented at the International Conference of Traffic and Transport Psychology, Valencia, Spain.*

Kinnear N, Hammond J, Smith L, Ramnath R, Simms G, Hitchings J and Helman S (2020). *The potential impact of Graduated Driver Licensing in Northern Ireland* (PPR934). Crowthorne: Transport Research Laboratory

Kinnear N, Lloyd L, Helman S, Husband P, Scoons J, Jones S, Stradling S, McKenna F and Broughton J (2013). *Novice Drivers: evidence review and evaluation: pre-driver education and training, graduated driver licensing, and the New Drivers Act* (PPR673). Commissioned and funded by the Department for Transport. Crowthorne: Transport Research Laboratory

Kinnear N, Kelly SW, Stradling S and Thomson JA (2013). Understanding how Drivers Learn to Anticipate Risk on the Road: A Laboratory Experiment of Affective Anticipation of Road Hazards. *Accident Analysis and Prevention*, 50, 1025-1033.

Kinnear N and Wallbank C (2020). *Graduated driver licensing in the Isle of Man* (PPR952). Crowthorne: Transport Research Laboratory.

Maycock G, Lockwood CR and Lester J (1991). *The accident liability of car drivers* (RR315). Crowthorne: Transport and Road Research Laboratory.

Mayhew DR, Simpson HM and Pak A (2003). Changes in collision rates among novice drivers during the first months of driving. *Accident Analysis and Prevention*, 35(5), 683-691.

Mayhew DR, Simpson HM and Robinson A (2002). The safety value of driver education and training. *Injury Prevention*, 8(Supp. II), ii3–ii8.

Mayhew DR, Simpson HM, Williams AF and Ferguson SA (1998). Effectiveness and role of driver education and training in a graduated licensing system. *Journal of Public Health Policy*, 19, 51–67.

McCartt AT, Shabanova VI and Leaf WA (2003). Driving experience, crashes and traffic citations of teenage beginning drivers. *Accident Analysis and Prevention*, 35, 311–320.

McCartt AT, Mayhew DR, Braitman KA, Ferguson SA and Simpson HM (2009). Effects of age and experience on young driver crashes: review of recent literature. *Traffic Injury Prevention*, 10(3), 209-219.

McCartt AT, Teoh ER, Fields M, Braitman KA and Hellinga LA (2010). Graduated licensing laws and fatal crashes of teenage drivers: a national study. *Traffic Injury Prevention*, 11, 240-248.

McKenna FP, Horswill MS and Alexander JL (2006). Does anticipation training affect drivers' risk taking? *Journal of Experimental Psychology: Applied*, 12(1), 1-10.

Montori VM and Guyatt GH (2001). Intention-to-treat principle. *Cmaj*, 165(10), 1339-1341.

O'Cathain A, Croot L, Duncan E, et al. (2019). Guidance on how to develop complex interventions to improve health and healthcare. *BMJ Open*; 9:e029954.

Ouimet MC, Pradhan AK, Brooks-Russell A, Ehsani JP, Berbiche D and Simons-Morton BG (2015). Young drivers and their passengers: a systematic review of epidemiological studies on crash risk. *Journal of Adolescent Health*, 57(1), S24-S35.

Quimby AR, Maycock G, Palmer C and Buttress S (1999). *The Factors that Influence a Driver's Choice of Speed: A Questionnaire Study* (TRL325). Crowthorne: Transport Research Laboratory.

Pressley A, Fernández-Medina K, Helman S, McKenna FP, Stradling S and Husband P (2016). *A review of interventions which seek to increase the safety of young and novice drivers* (PPR781). Commissioned and funded by the Department for Transport. Crowthorne: Transport Research Laboratory.

Roberts IG and Kwan I (2001). School-based driver education for the prevention of traffic crashes. *Cochrane Database of Systematic Reviews*, Issue 3. Art. No.: CD003201. https://doi.org/10.1002/14651858.CD003201

Russell KF, Vandermeer B and Hartling L (2011). Graduated driver licensing for reducing motor vehicle crashes among young drivers. *Cochrane Database of Systematic Reviews*, 10, CD003300.

Sagberg F (1998). Month-by-month changes in accident risk among novice drivers. *Paper presented at the 24th International Congress of Applied Psychology*, San Francisco, August 9–14.

Senserrick TM and Williams AF (2015). Summary of literature of the effective components of graduated driver licensing systems (No. AP-R476/15).

Sexton B and Grayson G (2010). *The accident history and behaviours of new drivers who pass their first practical driving test* (PPR427). Crowthorne: Transport Research Laboratory.

Smith JD, Li DH and Rafferty MR (2020). The Implementation Research Logic Model: a method for planning, executing, reporting, and synthesizing implementation projects. *Implementation Science*, 15, 84. https://doi.org/10.1186/s13012-020-01041-8

Soto CJ and John OP (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, 68, 69-81.

Tong S, Lloyd L, Durrell L, McCrae-Mckee K, Husband P, Delmonte E and Buttress S (2015). *Provision of telematics research* (PPR755). Crowthorne: Transport Research Laboratory.

Vaa T, Høye A and Almqvist R (2015). Graduated driver licensing: searching for the best composition of components. *Latin American Journal of Management for Sustainable Development*, 2(2), 160-176.

Vernick JS, Li G, Ogaitis S, MacKenzie EJ, Baker SP and Gielen AC (1999). Effects of high school driver education on motor vehicle crashes, violations and licensure. *American Journal of Preventative Medicine*, 16(1), 40–46.

VicRoads (2017). *Examination of the impact of the graduated licensing system on young novice driver safety: summary report.* Melbourne: VicRoads. ISBN 978-0-7311-9176-5

Weekley J and Helman S (2024). *Driver2020 – an evaluation of interventions designed to improve safety in the first year of driving. Supplementary appendix – data collection surveys* *and intervention logic models* (PPR2034). Commissioned and funded by the Department for Transport. Crowthorne: Transport Research Laboratory.

Weekley J, Helman S, Chowdhury S, Hammond J and Hutton J (2024a). *Driver2020 – an evaluation of interventions designed to improve safety in the first year of driving. Report D1: Effectiveness of interventions delivered to novice drivers* (PPR2009). Commissioned and funded by the Department for Transport. Crowthorne: Transport Research Laboratory.

Wegman F and Katrakazas C (2021). Did the COVID-19 pandemic influence traffic fatalities in 2020? A presentation of first findings. *IATSS research*, 45(4), 469-484.

Wells P, Tong S, Sexton B, Grayson G and Jones E (2008). Cohort II: a study of learner and new drivers. Volume 1: main report. Road Safety Research Report No. 81. London: Department for Transport.

West SG, Duan N, Pequegnat W, Gaist P, Des Jarlais DC, Holtgrave D, Szapocznik J, Fishbein M, Rapkin B, Clatts M and Mullen PD (2008). Alternatives to the randomized controlled trial. *American Journal of Public Health*, 98(8), 1359–1366. https://doi.org/10.2105/AJPH.2007.124446

Williams AF and Ferguson SA (2004). Driver education renaissance? *Injury Prevention*, 10 4–7.

Williams AF (2017). Graduated driver licensing (GDL) in the United States in 2016: A literature review and commentary. Journal of safety research, 63, 29-41.

Appendix A What is known about young and novice drivers – the starting point for the Driver2020 project

This section expands on the summary provided in Section 1.

A.1 Young and novice drivers represent a long-standing road safety challenge

Half a century ago, Goldstein (1972) noted "That youthful drivers…are over-represented in accidents…considerably beyond their proportion in the driving population, has been well known for several decades…". Within the same context Goldstein also drew attention to the widely observed fact that those with little experience of a new task tend to make more errors and show less dependable skill and judgement in its execution than those with more experience. Thus, it would not be a stretch to claim that the safety challenge presented by young and novice drivers is something that has been known for three-quarters of a century, and the group is still over-represented in fatal and serious crashes (e.g. House of Commons, 2021). Evidence from research into this group has confirmed that age and experience both play a role in an increased risk of being involved in road collisions.

A.2 Increased age and increased experience are associated with reduced risk

Studies in multiple countries have shown that the collision risk of new drivers is greater than that of more experienced drivers (Wells, Tong, Sexton, Grayson & Jones, 2008; Mayhew, Simpson & Pak, 2003; McCartt, Shabanova & Leaf, 2003; Sagberg, 1998; Forsyth, Maycock & Sexton, 1995; Maycock, Lockwood & Lester, 1991). These studies also show that the younger drivers are at a greater risk of collision than older drivers (although it should be noted that risk rises again in old age). Several reasons are offered as to what it is about younger age that leads to greater risk, including those associated with lifestyle (for example driving while under the influence of alcohol, and with friends in the car who distract the driver) and neuroscience (for example the underdevelopment of the frontal lobes – see Isler & Starkey, 2008).

Figure A-1 shows data reproduced from Maycock et al. (1991), from Great Britain. These data are modelled from self-reported collisions, with exposure kept constant. The dotted lines in the figure show (separately for males and females) the first-year accident liability for drivers passing their test at a given age. The solid lines show accident liability for people who pass their test (and therefore begin driving) at age 17, as they get older and accumulate on-road driving experience. Around 90% of the collisions in this dataset and others like it in UK studies are so-called 'damage only' collisions. The remaining 10% is dominated by collisions in which slight injuries occurred. Serious injury collisions tend not to be frequent enough to study statistically in this way in sampled datasets.

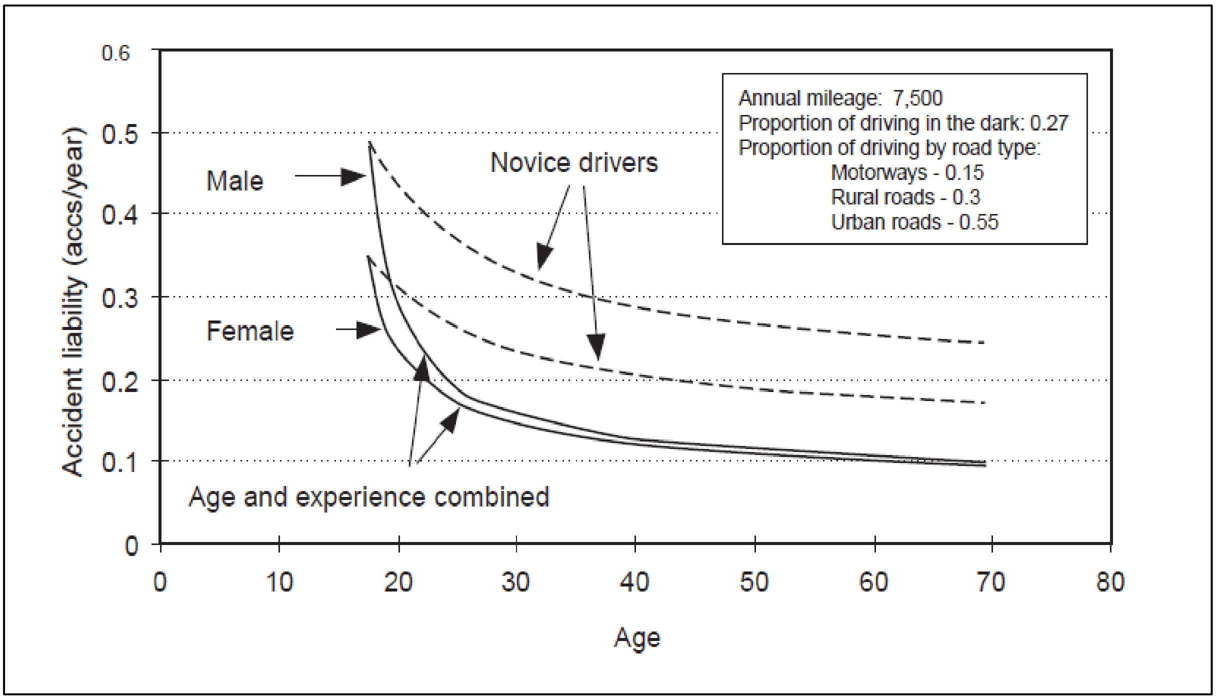


Figure A-1: The effects of age and experience on collisions risk (figure reproduced from Maycock et al., 1991)

While from data collected 30 years ago, this figure demonstrates the broad pattern of findings with respect to age and experience seen in modern datasets. All other things being equal, the younger a driver is when they begin unsupervised driving, the greater their risk of being involved in a collision, and new drivers of all ages become safer as they accumulate on-road experience. There are exceptions; for example, a subset of first-time passers in Sexton and Grayson (2010) were shown to have a lower crash risk despite being younger and reporting a driving style that has in general been associated with greater risk. However, the general protective effects of maturation and on-road experience are seen in multiple countries. Estimates suggest that most of the improvements in safety arising from experience come in the first 1,000-3,000 miles of independent driving (McCartt et al., 2003; see also Kinnear, Kelly, Stradling & Thomson, 2013).

It should be noted that collisions in the learner period, when drivers are supervised, are at a very low level compared with the levels reached when unsupervised driving begins (for example see Vicroads, 2017).

A.3 Approaches that target age and inexperience in higher risk situations have worked well to reduce risk in young and novice drivers

The most successful approaches to lowering the risk of collisions in this group have been those licensing approaches known collectively as ‘Graduated Driver Licensing’ (GDL). Such approaches focus explicitly on increasing the age at which drivers become licensed, and on increasing levels of on-road experience in safer conditions both before licensure (supervised driving), and afterwards (supervised driving, or solo driving in lower risk conditions). The evidence is summarised here; for readers interested in more detail, other reviews are available (Kinnear et al., 2020; Kinnear & Wallbank, 2020; Kinnear, Lloyd et al., 2013; Russell, Vandermeer & Hartling, 2011).

Age has been targeted in several ways in GDL systems. The first way is to simply define the age at which someone can begin to drive unaccompanied, and the evidence suggests that when this is set to an older age, this increases safety (McCartt, Mayhew, Braitman, Ferguson & Simpson 2009). A second way in which age is targeted is to increase the minimum time for which someone needs to remain in the 'learner' phase of the licensing system. Evaluations of this have consistently found that longer periods lead to greater safety benefits (Senserrick & Williams, 2015).

A longer learning period also gives more opportunity for on-road practice. This has also been targeted directly in some licensing systems through the setting of minimum requirements; evidence from Australia suggests that if at least 120 hours of on-road practice can be achieved, this probably has safety benefits (Senserrick & Williams, 2015), although the evidence on this issue is less well established than on increasing licensure age and the length of learner periods.

The approach to managing post-licence experience in GDL is built on an understanding of specific higher risk situations for young and novice drivers. The reasoning is that if experience can be allowed to build up initially either in low-risk conditions, or when supervised in higher risk conditions, the probability of collisions can be reduced. The two higher risk situations that are most well understood are driving when carrying peer-age passengers and driving at night.

For drivers aged under 25, carrying at least one passenger has been shown by a systematic review of the literature (Ouimet et al., 2015) to increase fatal crashes by 1.24 to 1.89 times relative to solo driving. The risk estimate for carrying two or more passengers was 1.70 to 2.92. It has been shown that passenger restrictions in GDL systems lead to overall reductions in risk (Senserrick & Williams, 2015; Williams, 2017; Vaa et al., 2015; Begg & Stephenson, 2003).

Younger drivers are also known to be over-represented in collisions at night in the UK (Clarke, Ward & Truman, 2002) and it is known that the more night-time driving is restricted in GDL systems in the US, the greater the reductions in collisions (McCartt et al., 2010).

There are other components that are often found in GDL systems. These include alcohol restrictions, speed limits, vehicle power restrictions and the use of vehicle identifiers to signal licensing status. The evidence on these has proven difficult to isolate from the effects of other GDL components (for a recent summary see Kinnear et al., 2020).

Overall, the evidence shows that GDL components that focus on age and on-road experience, both before and after licensing, are effective at reducing collision risk in novice drivers.

A.4 Approaches based on education, training, technology, and other mechanisms have not fared as well in reducing risk

The literature on interventions described using labels such as 'driver education' and 'driver training' is very large. Consequently, there have been numerous systematic and narrative reviews of it (Kinnear et al., 2013; Clinton & Lonero, 2006; Mayhew, Simpson & Robinson, 2002; Roberts & Kwan, 2001; Christie, 2001; Vernick, Li, Ogaitis, Mackenzie, Baker & Gielen, 1999; Mayhew, Simpson, Williams, & Ferguson, 1998; Brown, Groeger, & Biehl, 1987). These reviews have all come to the same conclusion, an example of which is the following quote from Helman, Grayson and Parkes (2010):

> "The only direct benefits imparted by broad driver education and training would appear to be the basic vehicle-control skills and knowledge of road rules necessary for entering the driving population. According to the evidence, it has no measurable direct effect on collision risk, and its continued use should therefore be set against much lower expectations in terms of what it can contribute directly to the safety of new drivers." (Helman et al., 2010, p8).

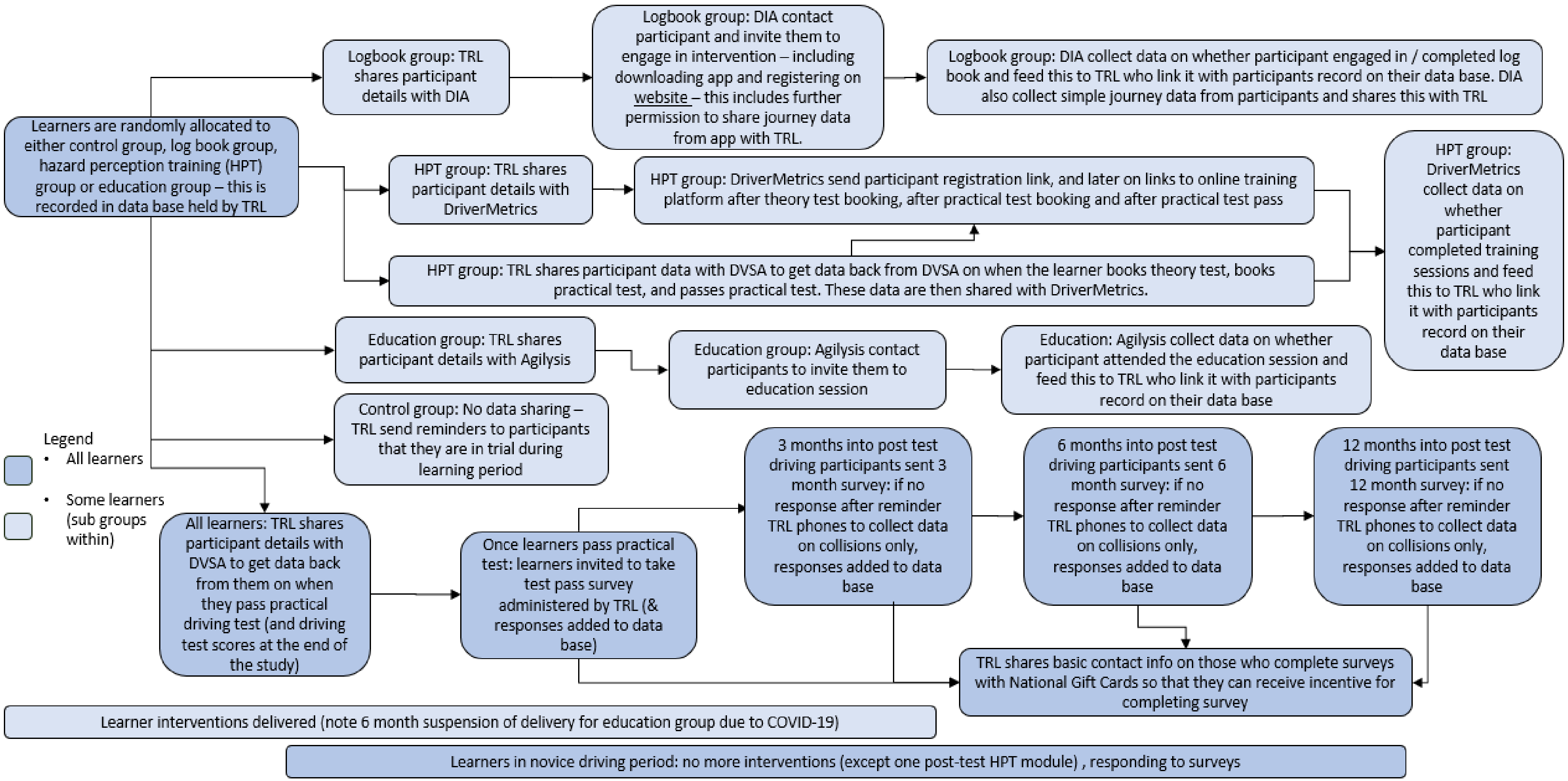
Such approaches can even cause harm; examples can be found in the literature regarding safety interventions such as skid-training courses that appear to promote over-confidence and risk-taking (Katila et al., 1996; Jones, 1993; Glad, 1988), and of driver education courses that can increase risk through allowing earlier licensure (Williams & Ferguson, 2004).

The conclusion from Helman et al. (2010) is necessarily in need of regular review since the specific theoretical approaches used in training and education (for example behaviour change techniques – see Fylan, 2017) and the technologies available to support them develop over time.

One promising approach, which illustrates the fast pace of technology development and the need for sound evaluation, is the use of 'telematics'. Typically, such technologies are used as part of insurance policies for young drivers. Tong et al. (2015) reviewed the literature on such approaches and concluded that no convincing evidence existed that they reduce risk. According to Tong et al., one of the main issues with the literature is that it is not feasible to utilise insurance data alone in evaluating such technologies, because of the self-selection and insurer-selection biases inherent in such datasets. Without properly matched control groups of drivers who do not have such policies, it is not possible to properly evaluate any change in risk associated with having such a policy independently of other effects, such as the types of drivers who typically have such policies.

One study that does have data on telematics policies, with a comparison group – albeit not one free from the biases mentioned above – is that of Helman et al. (2017). In a dataset from over 4,000 novice drivers in the UK, the study examined those factors associated with collision risk at six months post-test, when age and exposure were controlled. Those drivers in the sample with a telematics-based insurance policy reported *more* collisions at six months post-test than those without such a policy (the increase in risk after correcting for exposure and age was 50%). This concerning finding may have arisen due to biases in the groups and illustrates the importance of undertaking research from which causality can be inferred; innovation without controlled evaluation is not enough.

Appendix B Learner interventions – data flows



Appendix C Statistical testing

Statistical tests were used throughout this study to test for differences between groups (for instance, control, logbook, hazard perception training or education course groups, or between pre-pandemic and post-pandemic groups).

In all cases where statistical tests were used to compare data, the convention from the behavioural sciences of reporting *p*-values was adopted. The probability value, *p*-value, is used to determine statistical significance in a hypothesis test. If the probability of obtaining the value of the test statistic is less than 0.05, then the null hypothesis (or the assumption that the groups are equal on the same measure) is rejected and the result is termed 'statistically significant' (less than 5% likely to have occurred by chance alone). Tests where $p<0.05$ are therefore considered significant, however exact *p*-values have been reported throughout except where $p<0.01$.

Statistical significance alone can be misleading in trials with very large samples because it is influenced by the sample size. Increasing the sample size makes it more likely to find a statistically significant result, even for very small effects. As the sample size increases, the size of the effect that can be detected on a given measure decreases (i.e., one can identify very small or minor changes in behaviour). Therefore, this study presents the effect sizes where the results were statistically significant. Effect sizes help understand the magnitude of differences found. Both are essential to understand the full impact of the interventions.

A number of different tests were used in this study, depending on the type of data collected. Depending on the type of data, the tests could assume independence (responses are obtained from different individuals) or repeated measures analysis (the same individual responded multiple times). For instance, tests looking at the differences over the three reporting periods were repeated measures tests, whereas tests comparing the overall differences between the control and intervention groups assumed independence. Both parametric and non-parametric approaches were used based on the satisfaction (or not) of the parametric test assumptions; parametric approaches are more powerful but require the data to meet a number of criteria before they can be used.

The meaning of an effect size varies on the type of statistical test that is being used; the same effect size value has different meanings for different statistical tests. To give an indication of real-world impact, values of a given effect size statistic are described as 'negligible', 'very small', 'small', 'medium' or 'large'.

Techniques used include:

- Chi-square tests for independence used to compare the relationship between categorical variables. This is a non-parametric test. The effect size was determined using Cramer's V. Chi-square tests were used extensively during the analysis, for example to compare demographic distributions between groups in sections 4.3.1 and 4.5.2.
- Independent sample t-tests to compare mean values of a continuous variable for two groups or the non-parametric alternative, Mann-Whitney U tests. Instead of comparing means like the t-test, the Mann-Whitney test converts the scores into ranks and compares the median of the two groups. Effect sizes for the Mann-Whitney U tests were determined using Cliff's Delta. These tests were used in section 4.5.3 as

part of the per protocol analysis. Mann Whitney U tests were also used to compare the distribution of the personality variables in section 4.5.2.

- Between-participants analysis of variance (ANOVA) to compare the change in mean values of a continuous variable across an independent variable with more than two 'levels' (for example, intervention group). Generalised eta square was used to determine the effect size. These tests were used in section 4.5.3 as part of the per protocol analysis to assess differences between groups at each time point. The non-parametric alternative, Kruskal Wallis, was used if the assumptions of the parametric test were not satisfied; the effect size in this case was estimated as the eta squared based on the H-statistic. Kruskal Wallis tests were used in section 4.3.4 to assess differences between groups in learning and driving conditions.
- Repeated measures analysis of variance (ANOVA) to compare the change in mean values of a continuous variable over time (and across independent variables). Generalised eta square was used to determine the effect size. These tests were used in section 4.5.3 as part of the per protocol analysis to assess differences over time within the same group. The non-parametric alternative, Friedman test, was used if the assumptions of the parametric test were not satisfied, with effect sizes estimated using Kendall's W value. Freidman tests were also used in section 4.5.3.

C.1 Factor analysis

In addition to the tests described above, factor analysis was used to reduce some of the behaviour and attitudinal questions from a large number of related variables into a smaller number of linearly independent variables, reflecting the same underlying information. Based on common practice for these scales, factor analysis was used to combine multiple high correlated items into a single variable (factor). The output from the factor analysis for each question was the number of factors and the items relating to each factor – for example for the 'attitudes to enforcement and restrictions' (eight items combined into three factors, one item was removed) and 'driving style' (12 items combined into three factors, one of the items partially correlated with two of the factors).

The number of factors was determined using the 'scree test'. This involves looking at a plot of eigenvalues for factors and determining where the shape of the curve changes. For both sets of items, a three-factor solution was identified.

To determine the items relating to each factor, the factor loadings were used. The loadings (scores between -1 and 1) indicate the strength and direction of the association between each factor and item. Only one item ('Considerate – Selfish' in the driving style items) related strongly to more than one factor. Scores for each factor were then calculated using the coefficients from the factor analysis.

The factor analysis used data from the test pass survey only to determine the number of factors, the survey items and the coefficients relating to each factor. These were then used to calculate the factor scores for participants across all surveys. This ensured that the results are presented using the same factor coefficients and any statistical differences (analysed using the tests described above) can be attributed to a change over time.

Appendix D Characteristics of engagers and non-engagers

As discussed in section 4.5, as part of the per-protocol analysis that was carried out, the characteristics of the engagers for each intervention group were compared with the characteristics of the non-engagers to determine if the 'type' of person who engaged differed from the 'type' who did not. Where differences were identified for a characteristic variable, the results were shown in the main body of this report. This appendix shows the full results of all the comparisons carried out where no differences were identified.

D.1 Logbook intervention

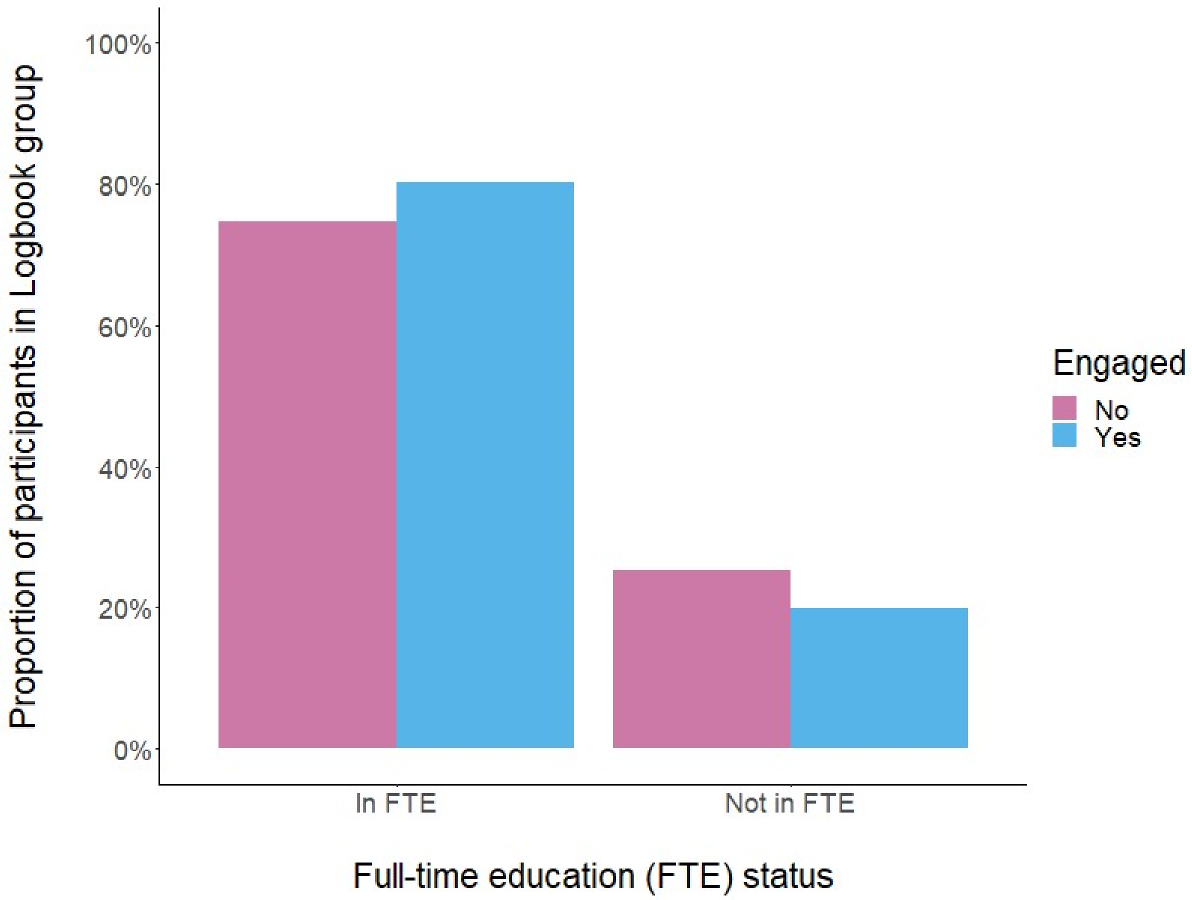


Figure D-1: Proportion of engagers and non-engagers in the logbook group by education status (N=4,438)

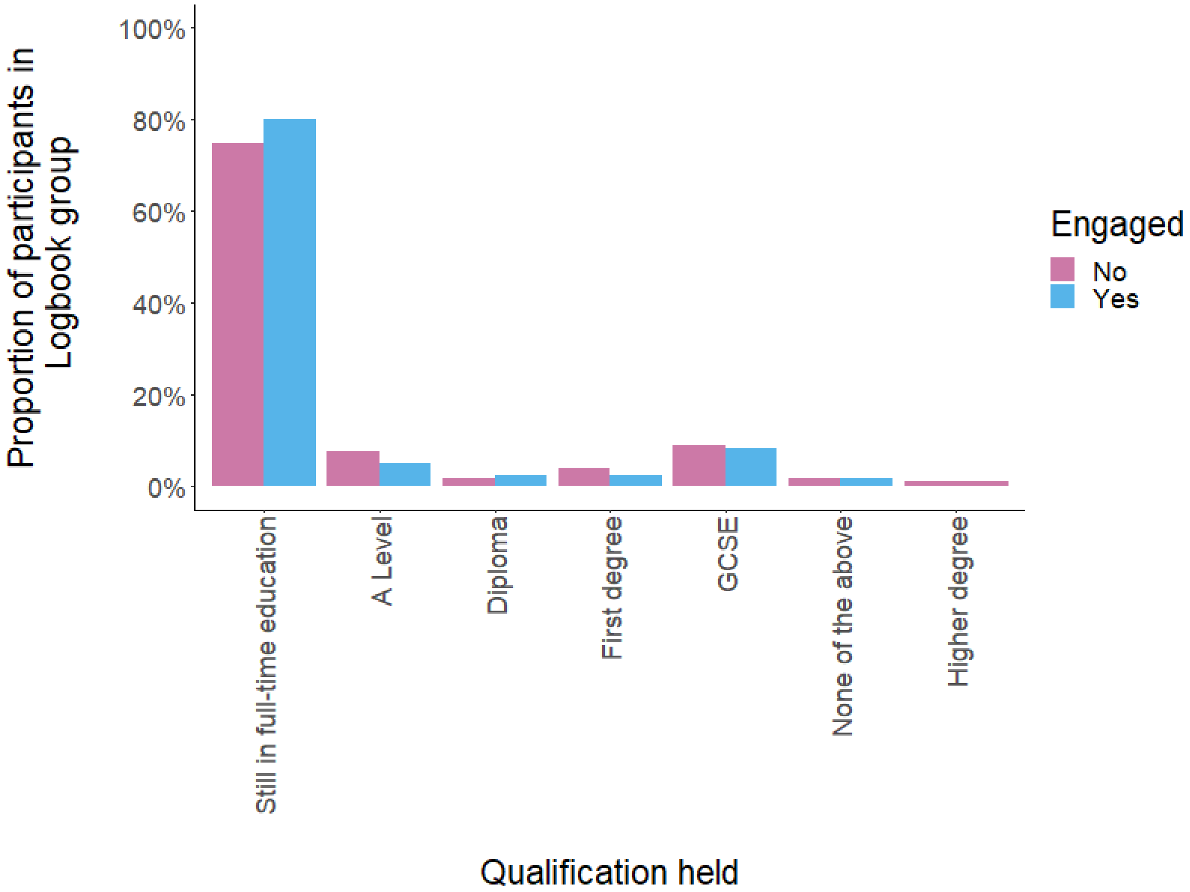


Figure D-2: Proportion of engagers and non-engagers in the logbook group by highest education qualification held (N=4,438)

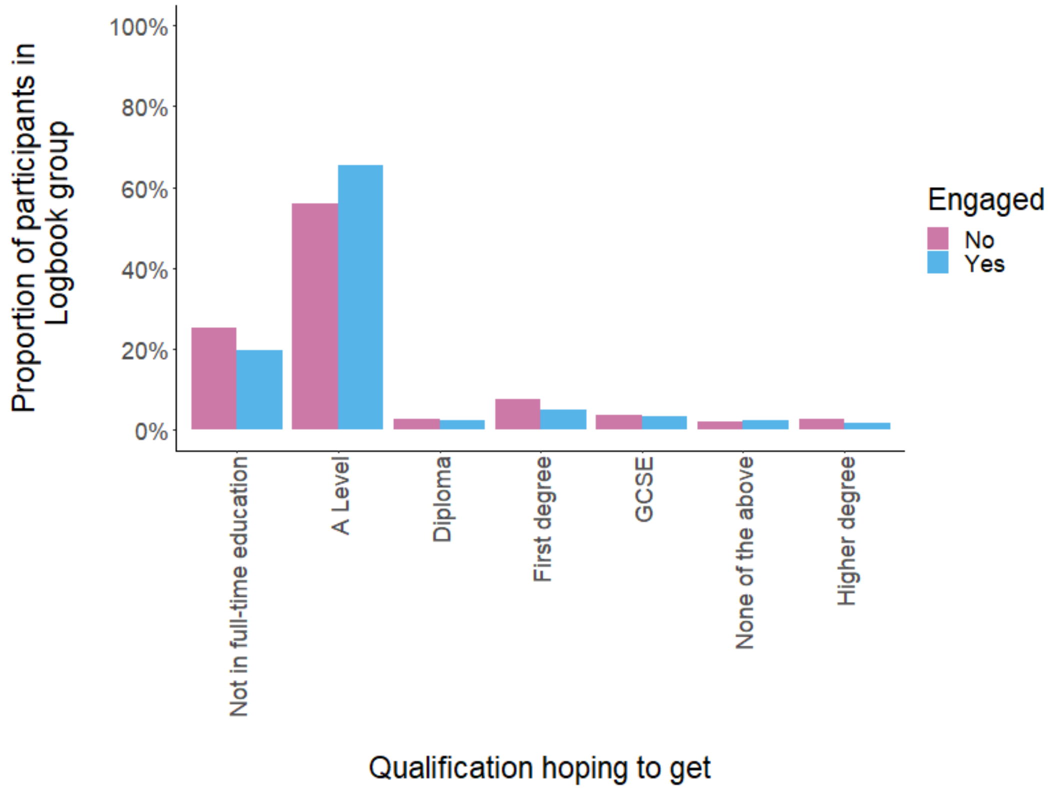


Figure D-3: Proportion of engagers and non-engagers in the logbook group by education qualification hoping to get (N=4,438)

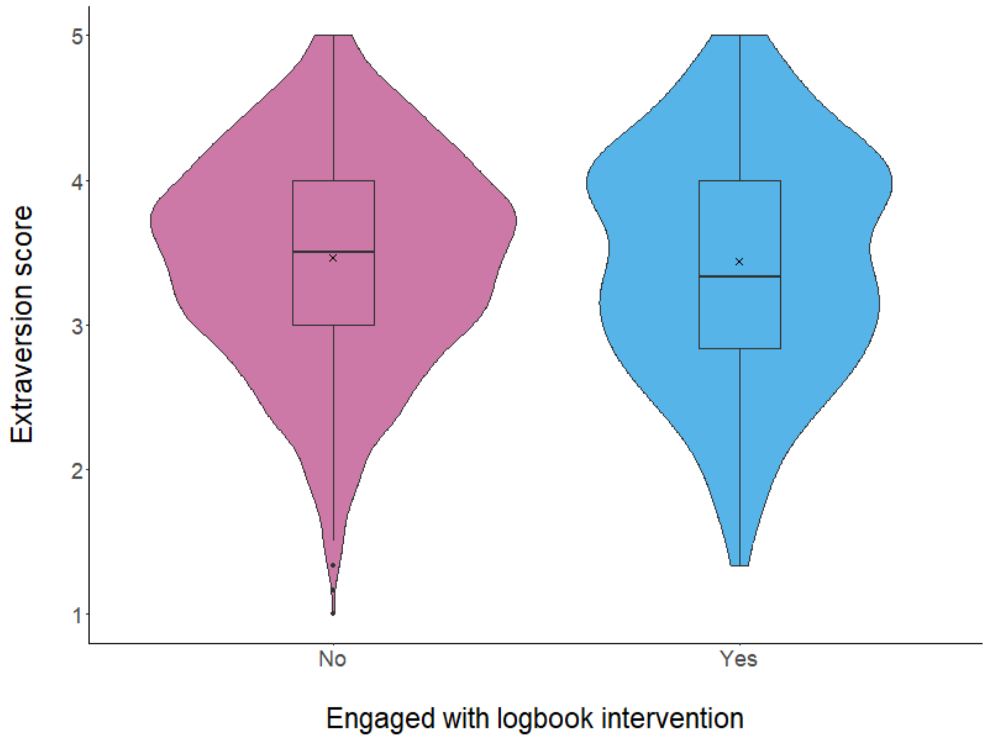


Figure D-4: Distribution of extraversion score by engagers and non-engagers in the logbook group (1=least, 5=most) (N=4,438)

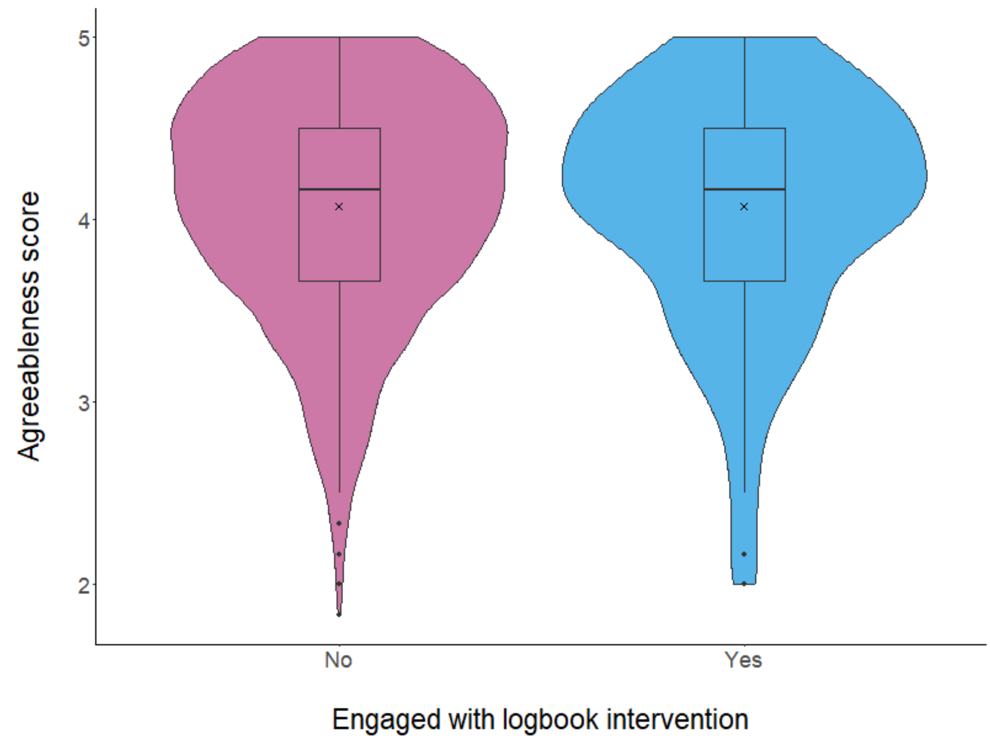


Figure D-5: Distribution of agreeableness score by engagers and non-engagers in the logbook group (1=least, 5=most) (N=4,438)

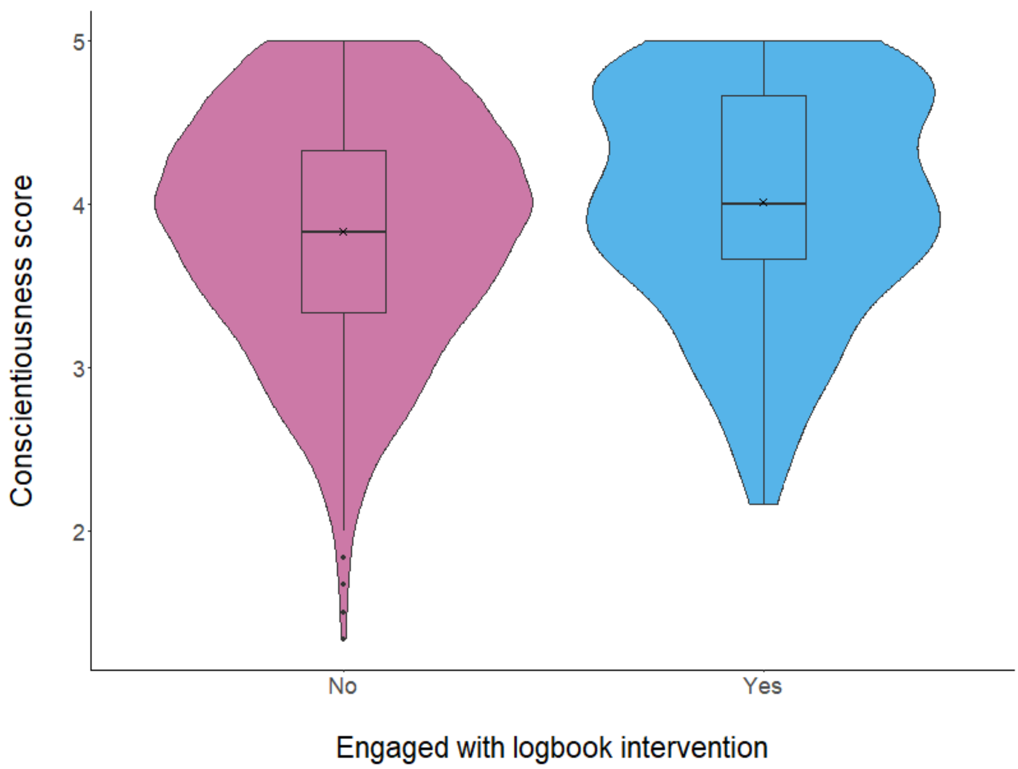


Figure D-6: Distribution of conscientiousness score by engagers and non-engagers in the logbook group (1=least, 5=most) (N=4,438)

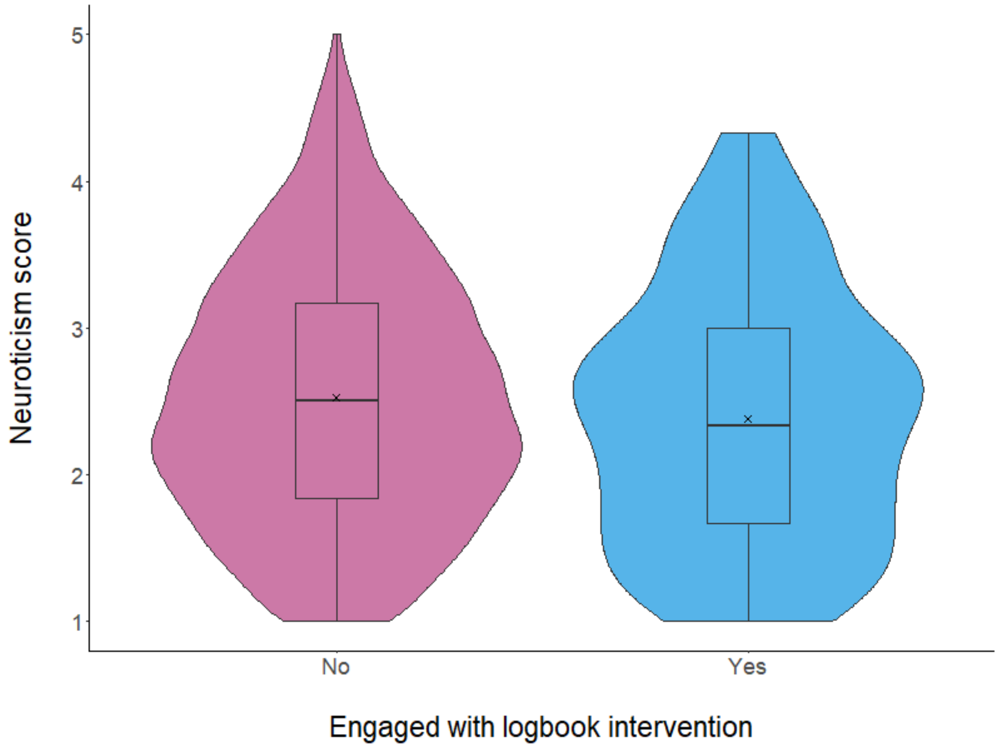


Figure D-7: Distribution of neuroticism score by engagers and non-engagers in the logbook group (1=least, 5=most) (N=4,438)

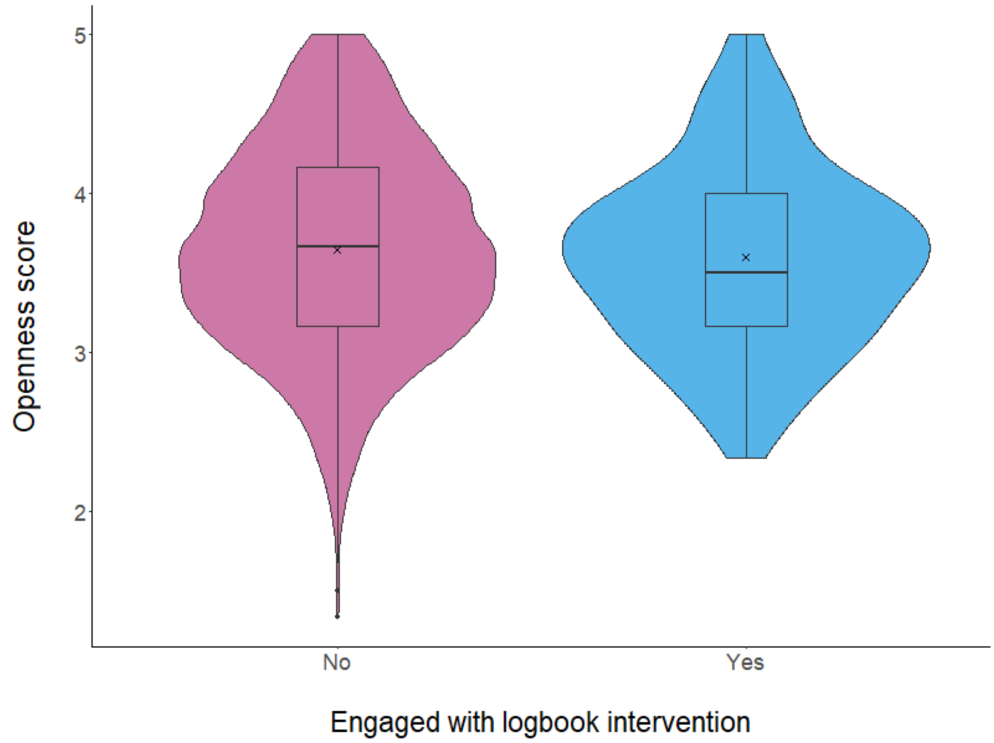


Figure D-8: Distribution of openness score by engagers and non-engagers in the logbook group (1=least, 5=most) (N=4,438)

D.2 Hazard perception training intervention

Figure D-9: Proportion of engagers and non-engagers in the hazard perception training group by gender (N=3,851)

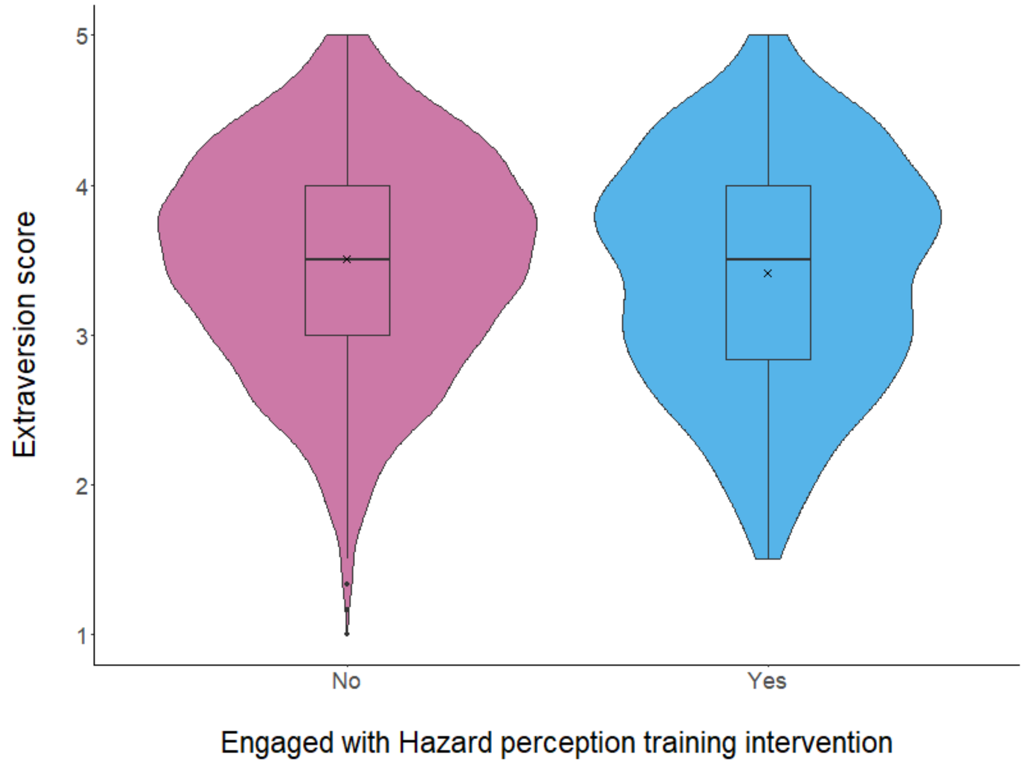


Figure D-10: Distribution of extraversion score by engagers and non-engagers in the hazard perception training group (1=least, 5=most) (N=3,851)

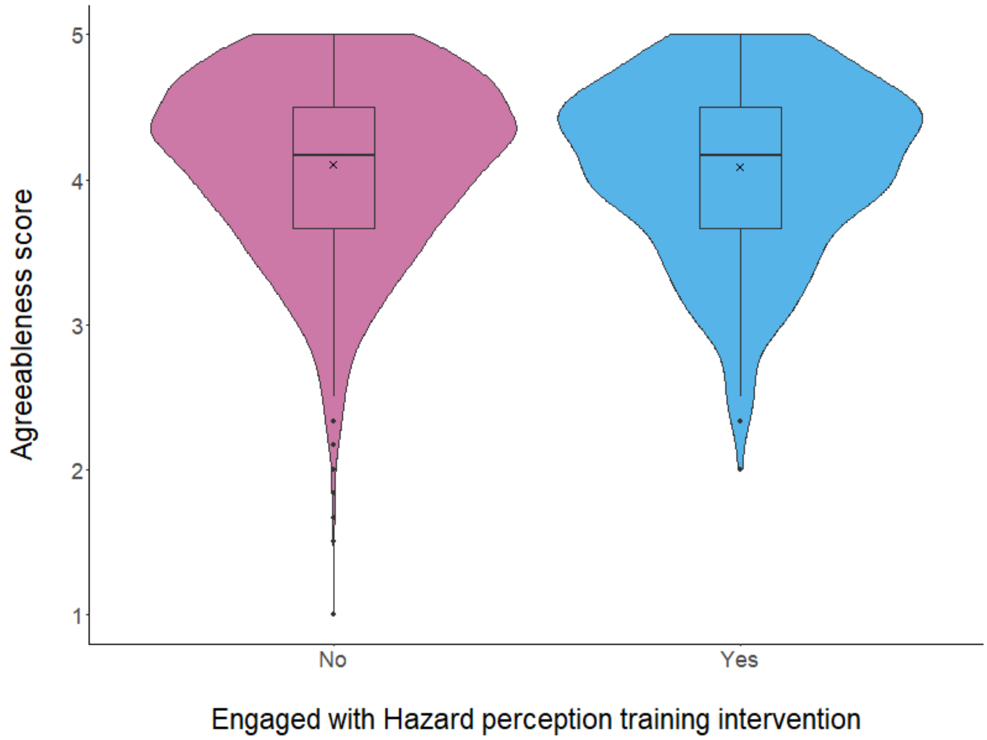


Figure D-11: Distribution of agreeableness score by engagers and non-engagers in the hazard perception training group (1=least, 5=most) (N=3,851)

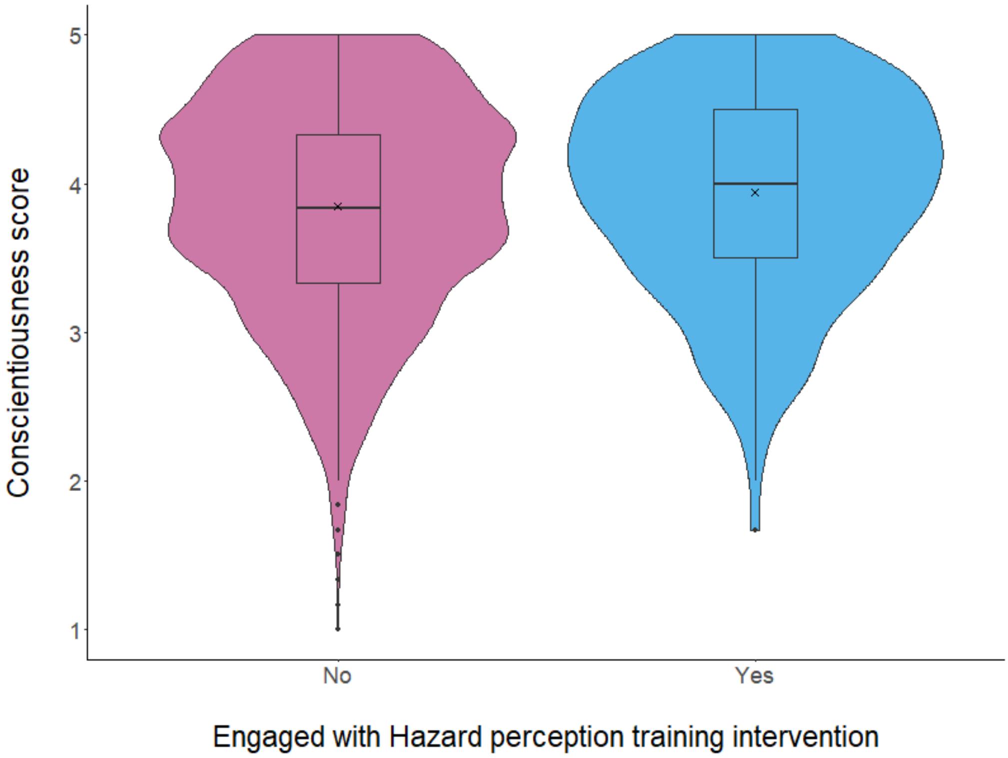


Figure D-12: Distribution of conscientiousness score by engagers and non-engagers in the hazard perception training group (1=least, 5=most) (N=3,851)

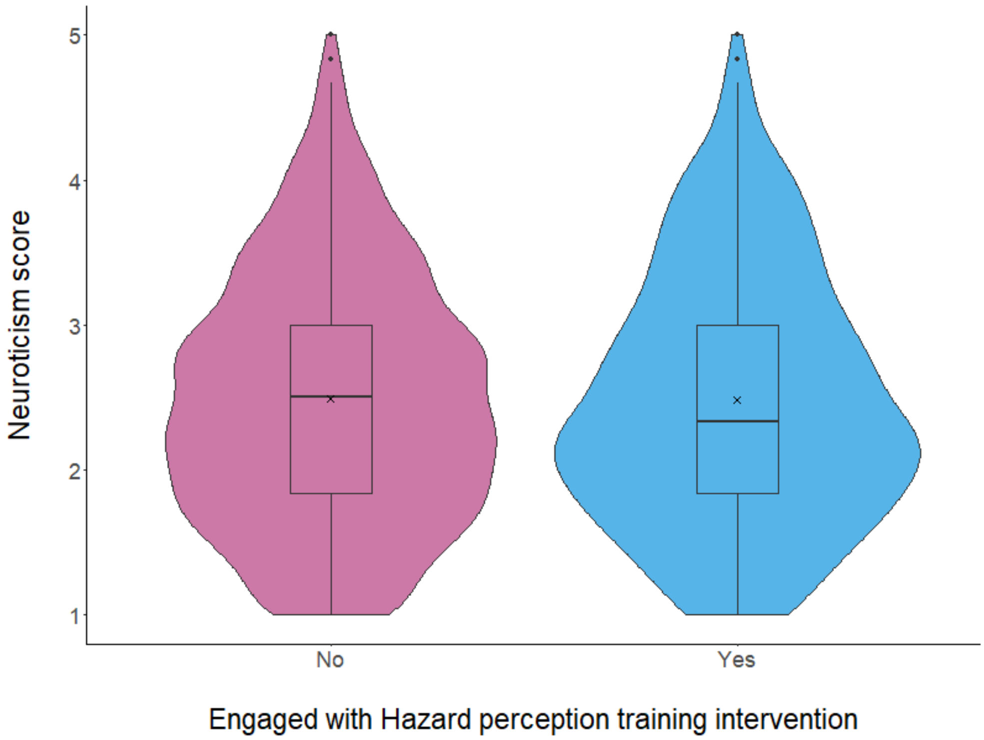


Figure D-13: Distribution of neuroticism score by engagers and non-engagers in the hazard perception training group (1=least, 5=most) (N=3,851)

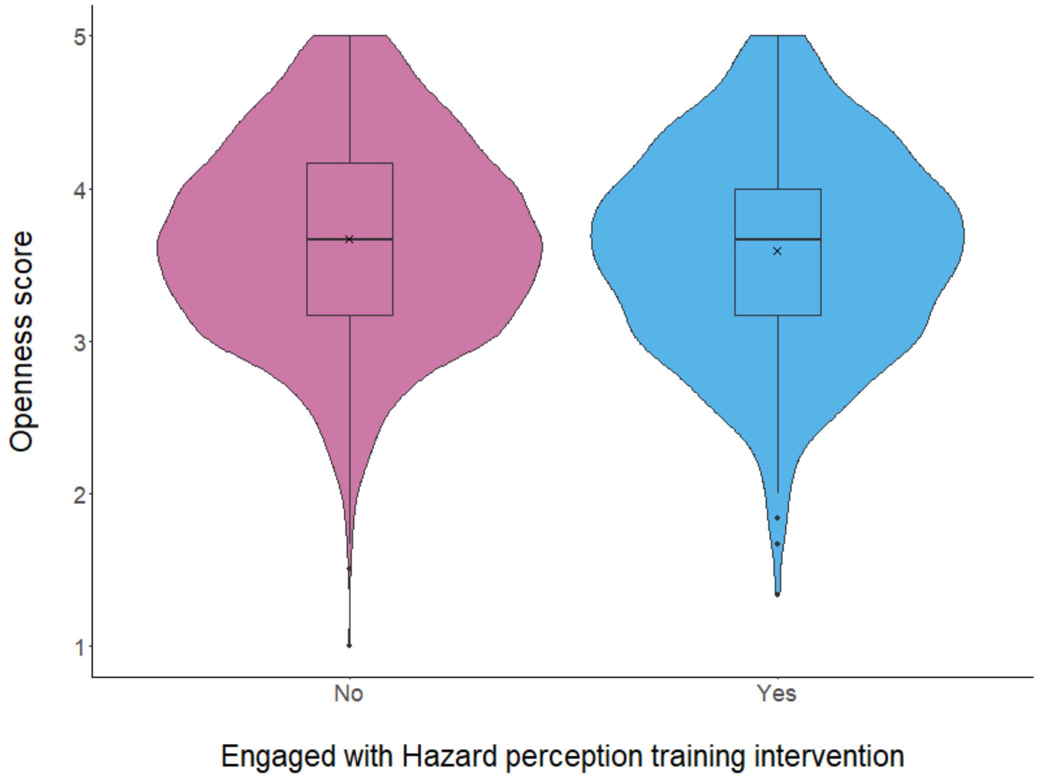


Figure D-14: Distribution of openness score by engagers and non-engagers in the hazard perception training group (1=least, 5=most) (N=3,851)

D.3 Education course intervention

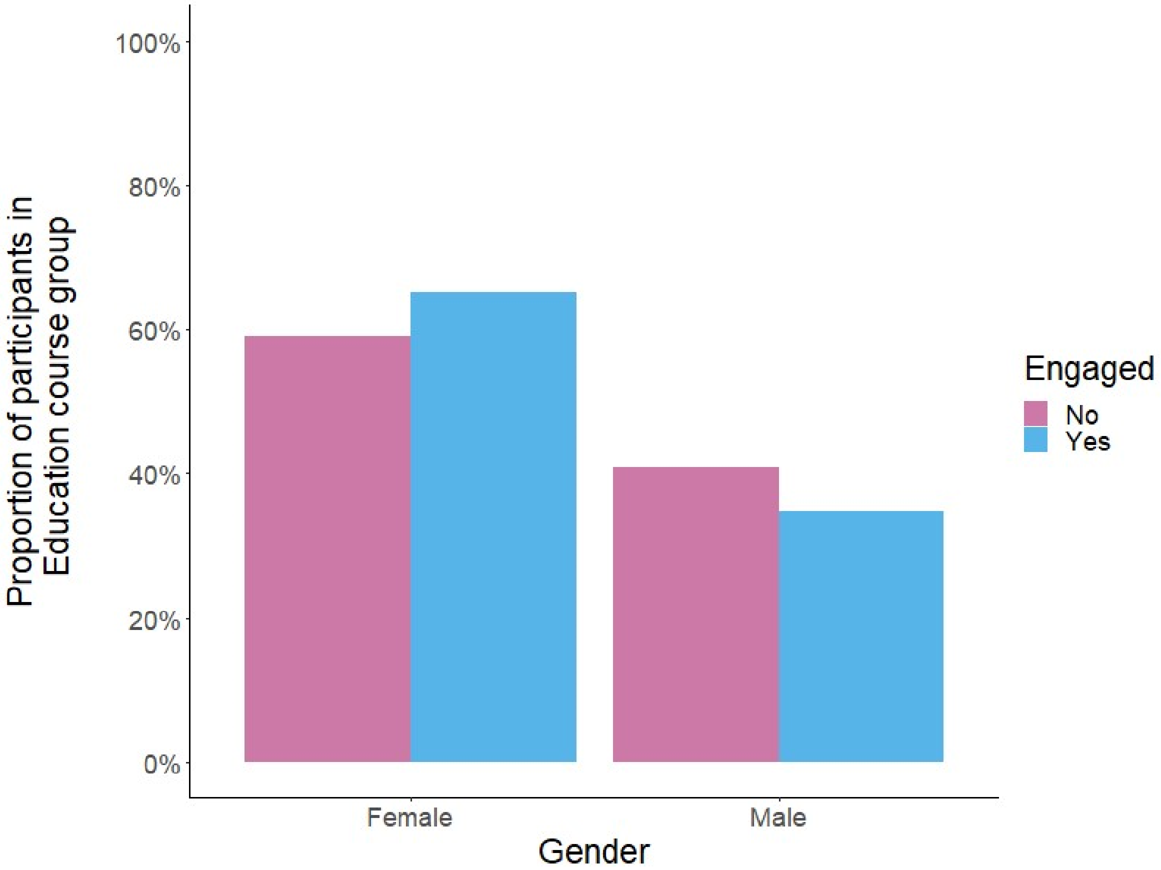


Figure D-15: Proportion of engagers and non-engagers in the education course group by gender (N=3,850)

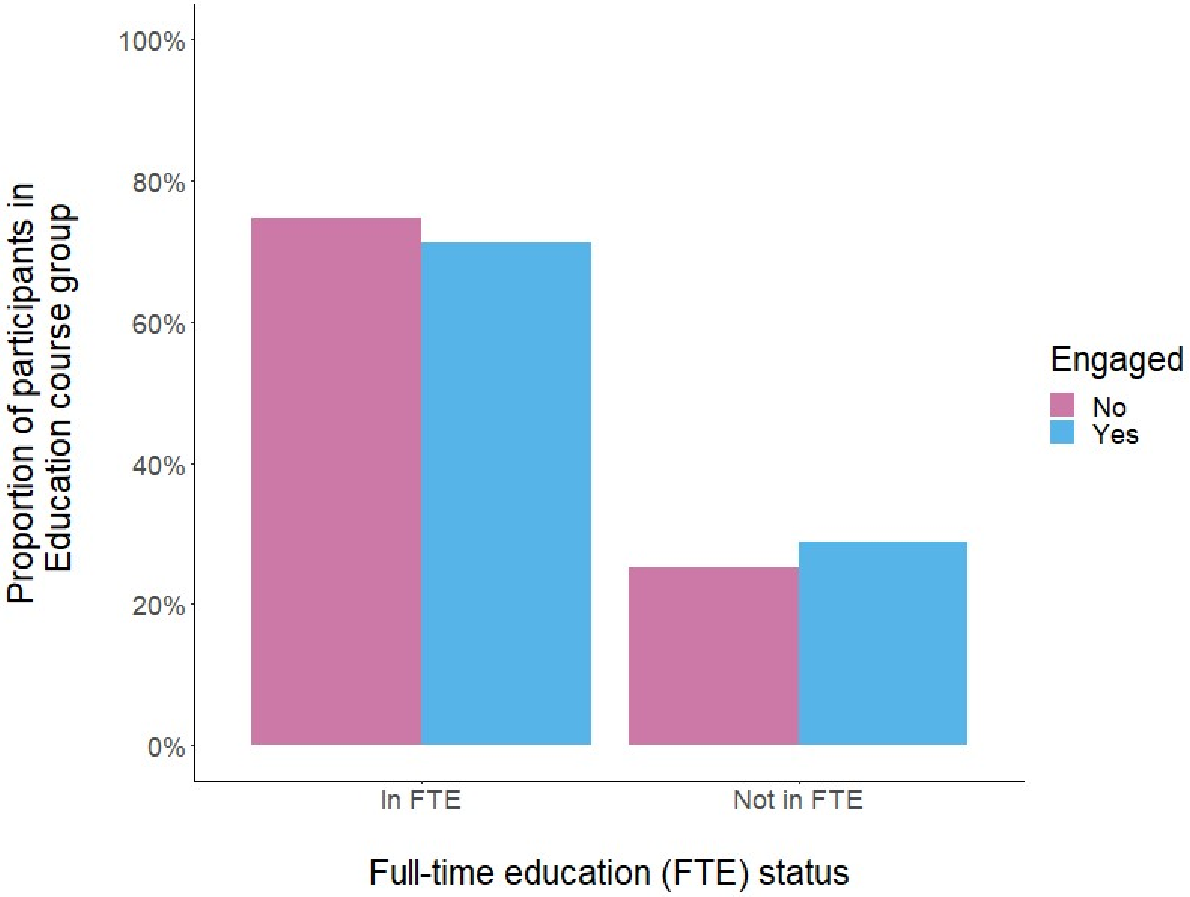


Figure D-16: Proportion of engagers and non-engagers in the education course group by education status (N=3,850)

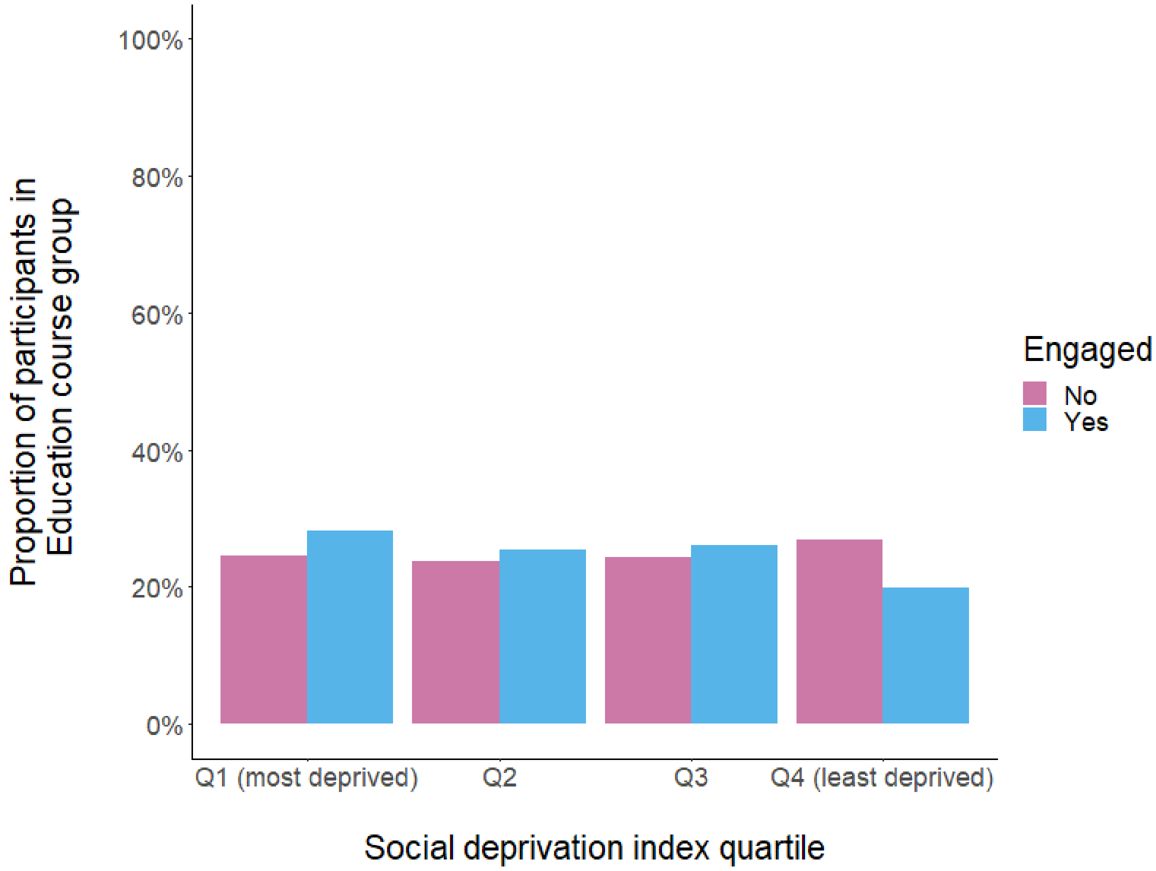


Figure D-17: Proportion of engagers and non-engagers in the education course group by SDI quartile (N=3,850)

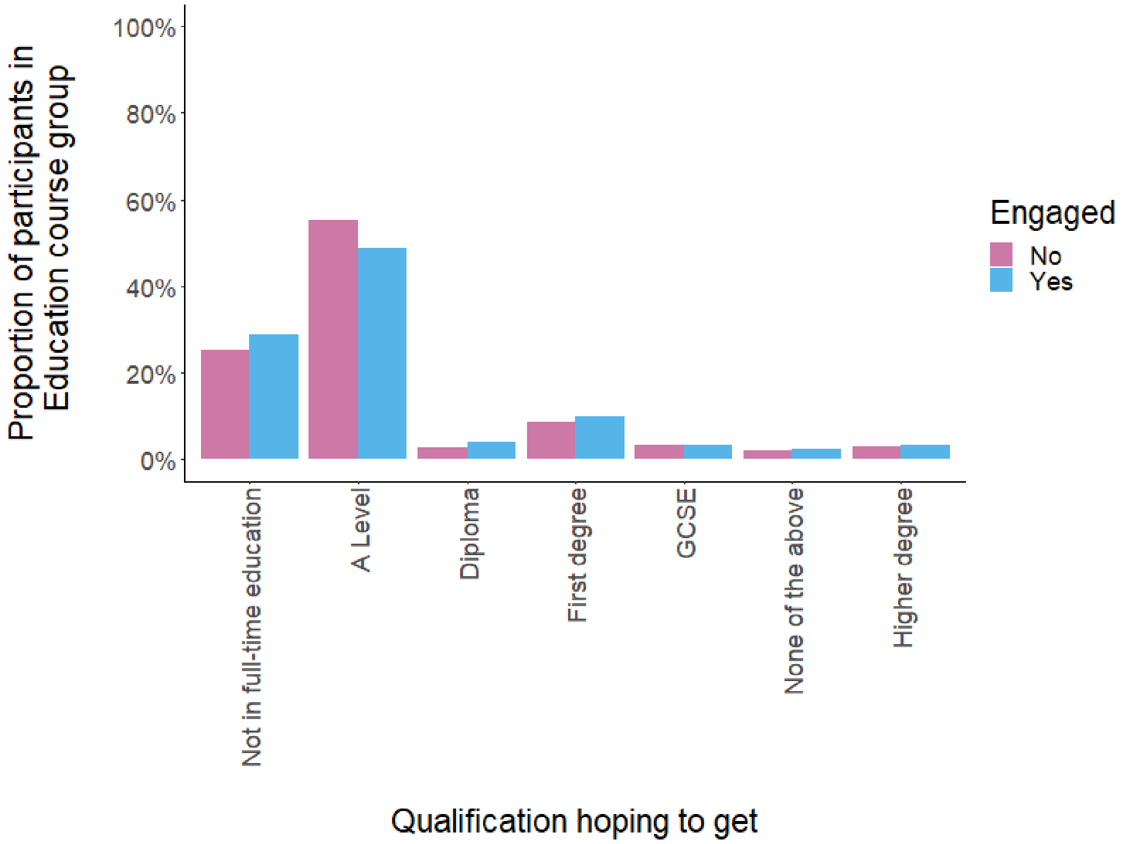


Figure D-18: Proportion of engagers and non-engagers in the education course group by education qualification hoping to get (N=3,850)

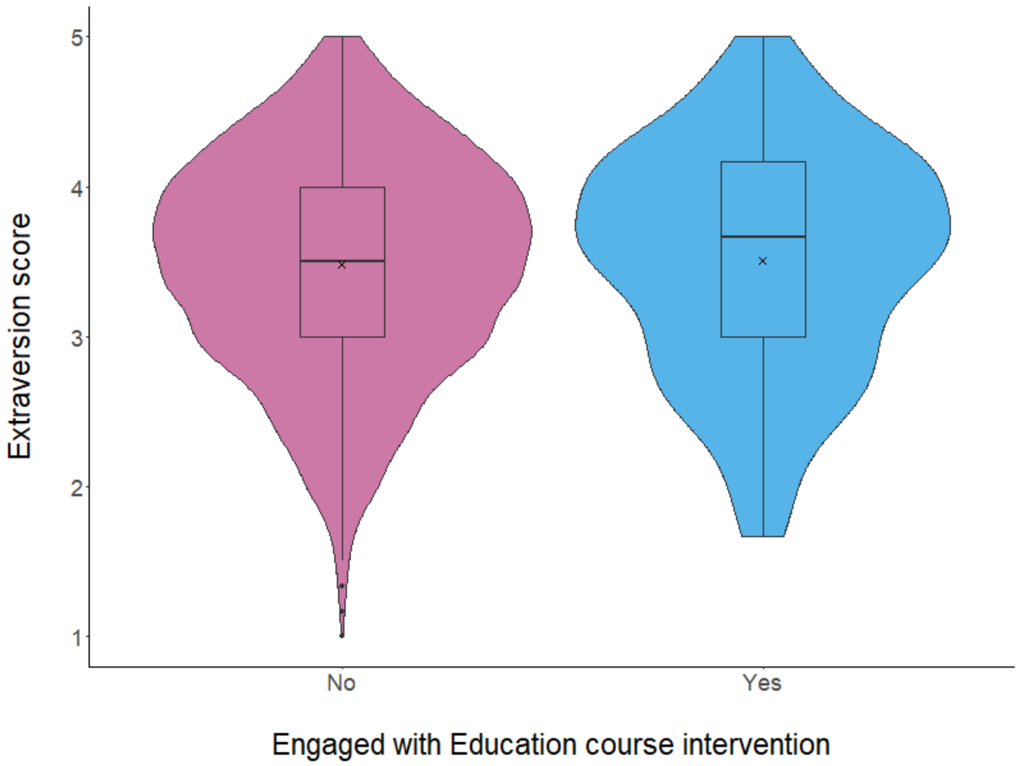


Figure D-19: Distribution of extraversion score by engagers and non-engagers in the education course group (1=least, 5=most) (N=3,850)

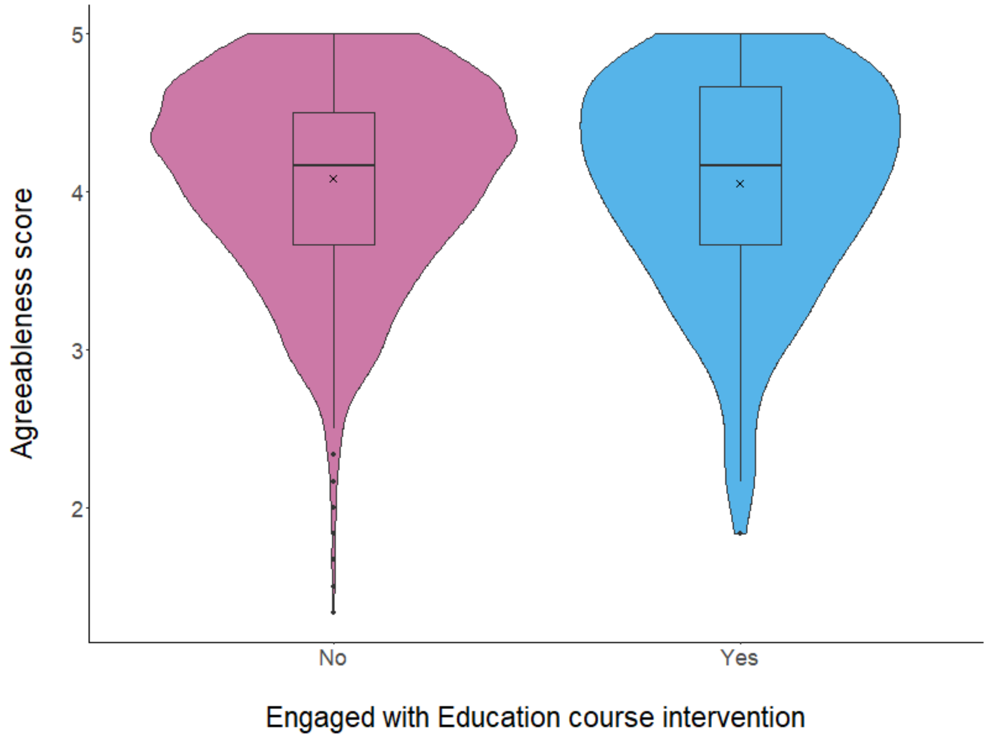


Figure D-20: Distribution of agreeableness score by engagers and non-engagers in the education course group (1=least, 5=most) (N=3,850)

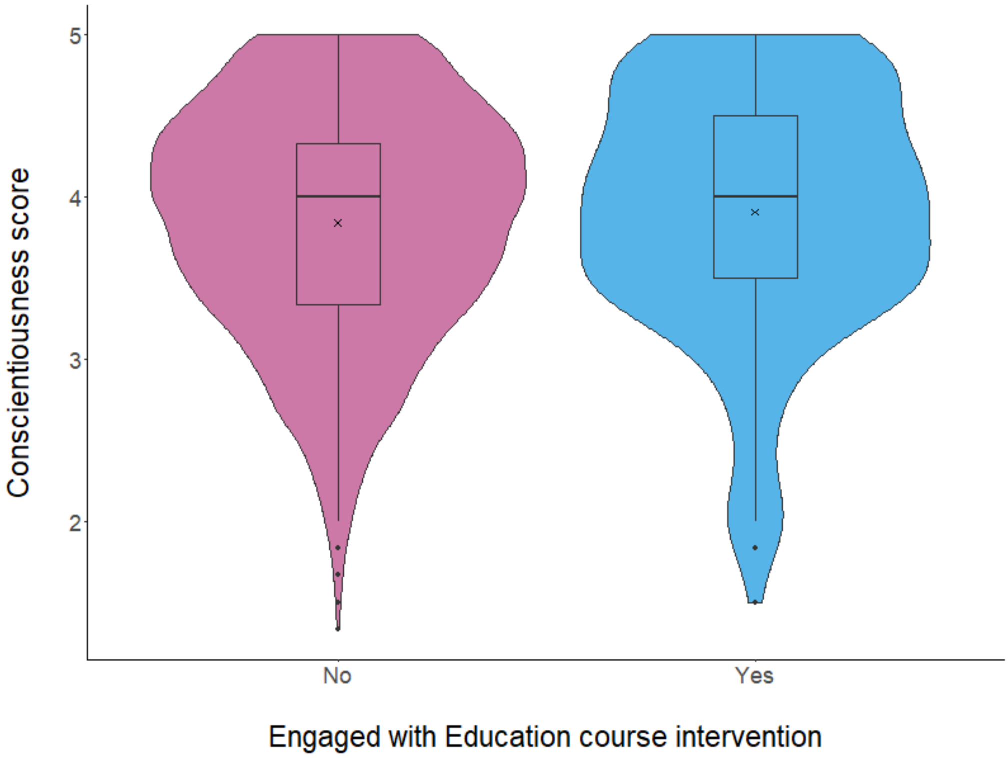


Figure D-21: Distribution of conscientiousness score by engagers and non-engagers in the education course group (1=least, 5=most) (N=3,850)

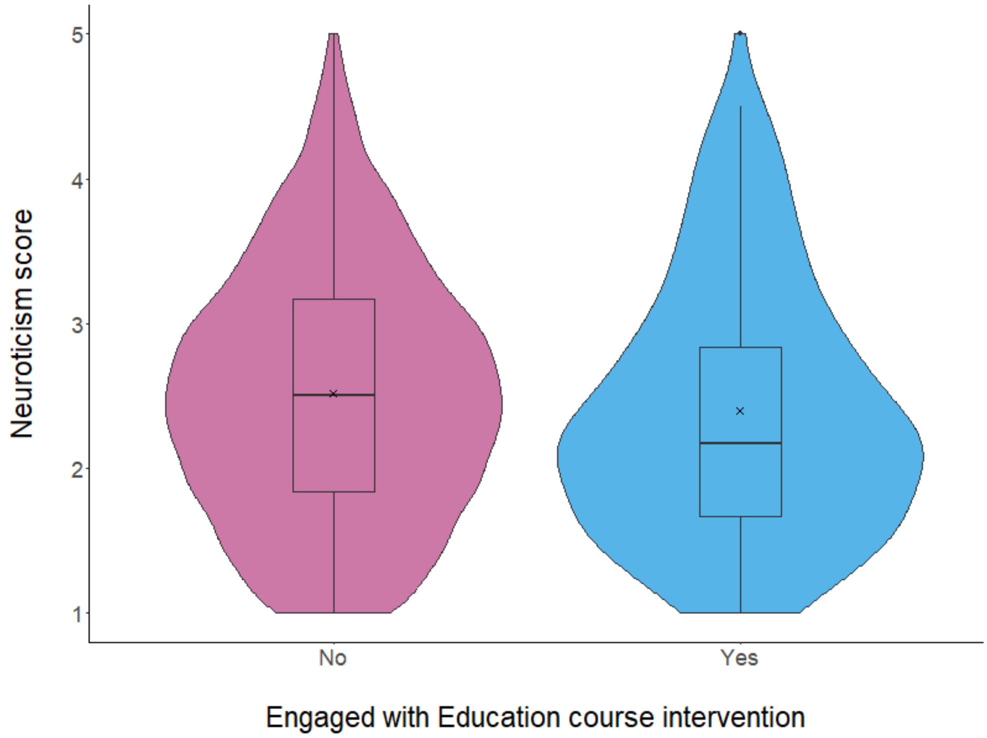


Figure D-22: Distribution of neuroticism score by engagers and non-engagers in the education course group (1=least, 5=most) (N=3,850)

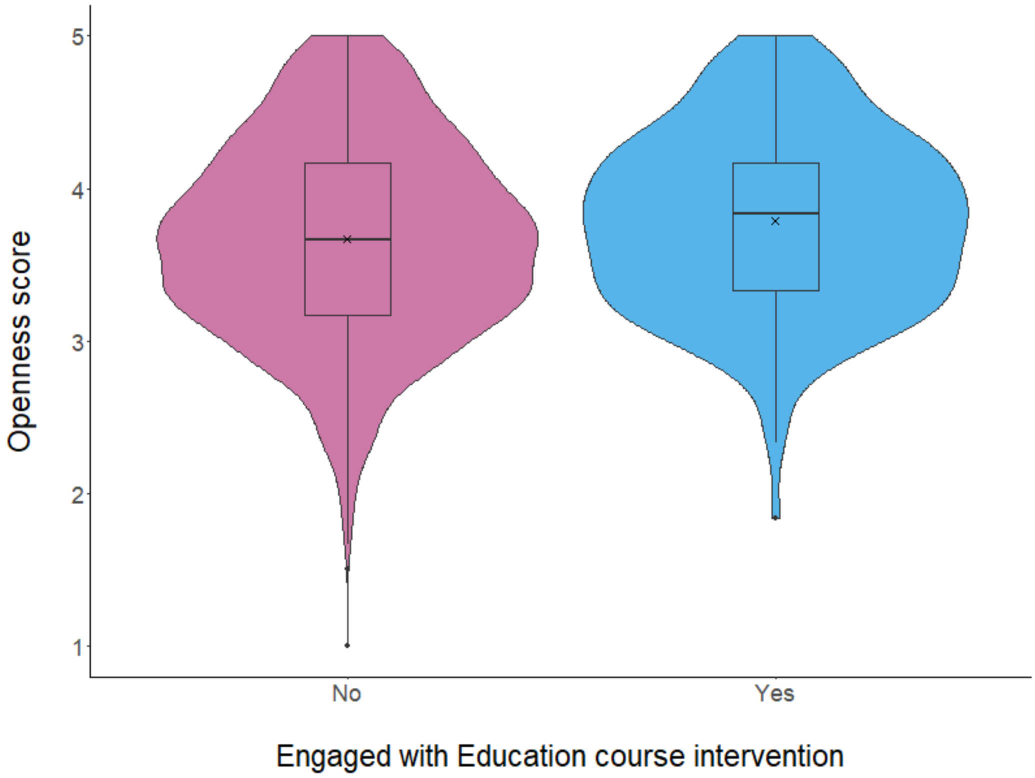


Figure D-23: Distribution of openness score by engagers and non-engagers in the education course group (1=least, 5=most) (N=3,850)

Appendix E Per-protocol analysis – matching process

E.1 Logbook intervention

Participants were matched on the three variables found to be different for engagers and non-engagers: age, gender, and SDI quartile. These three variables were used to sort the engagers into 72 ‘cells’, defined as the combinations of the different options for each of the variables (for example, male 17-year-olds in quartile 1). The control group participants were also sorted into the same cells.

To create the control group sample for use in the per-protocol analysis, eight control group participants were randomly chosen from each cell to correspond with each engager in that cell. So, for example, for every engager who was a male 17-year-old in SDI quartile 1, eight male 17 year olds in SDI quartile 1 were randomly selected from the control group and were included in the control group sample. A ratio of 1:8 was used as it provided the largest possible control group sample that was still very well matched to the logbook engagers.

This process resulted in a control group sample of 961 participants with age, gender, and SDI quartile characteristics which matched those of the participants who engaged with the logbook intervention. There was one cell where there was an insufficient number of control group participants in that cell to achieve the desired 1:8 ratio of engagers to control participants; hence the final sample of control group participants is 961, rather than 968 (121 x 8).

E.2 Hazard perception training intervention

Matching on all five variables where significant differences were identified would have created too many ‘cells’, reducing the sample size of the matched sample considerably. Therefore, different sampling methods were tested, matching on different combinations of the five variables. By matching on SDI quartile, highest qualification achieved and qualification hoping to get, the imbalances in all five of the variables were addressed. There were no statistically significant differences in any of the five variables between the matched control group sample and the hazard perception training engagers.

In more detail, the matching process was similar to that described above for logbook engagers:

- Using SDI quartile, highest qualification achieved and qualification hoping to get, the engagers were sorted into 196 ‘cells’, defined as the combinations of the different options for each of the variables (for example, participants in Q1, with highest qualification GCSEs and hoping to get A levels). The control group participants were also sorted into the same cells.
- To create the control group sample for use in the per-protocol analysis, three control group participants were randomly chosen from each cell to correspond with each engager in that cell. So, for example, for every engager who was in Q1, with highest qualification GCSEs and hoping to get A levels, three participants in Q1, with highest qualification GCSEs and hoping to get their A levels were randomly selected from the control group and included in the control group sample. A ratio of 1:3 was chosen as

a balance of maintaining control group sample size and obtaining a well-matched sample with no statistically significant differences.

- This process resulted in a control group sample of 1,220 participants with SDI quartile, highest qualification achieved and qualification hoping to get characteristics which matched those of the participants who engaged with the hazard perception training intervention. Furthermore, the control group participants were not statistically different from the engagers in age or education status characteristics after this matching was undertaken. There was one cell where there was an insufficient number of control group participants in that cell to achieve the desired 1:3 ratio of engagers to control participants. This is why there are 1,220 control group participants, rather than 1,236 (412 x 3).

E.3 Education intervention

Participants were matched on the two variables that had been found to be different for engagers and non-engagers: age and highest qualification held. These two variables were used to sort the engagers into 63 ‘cells’, defined as the combinations of the different options for the two variables (for example, 17-year-olds still in full time education). The control group participants were also sorted into the same cells.

To create the control group sample for use in the per-protocol analysis, six control group participants were randomly chosen from each cell to correspond with each engager in that cell. So, for example, for every engager who was a 17-year-old still in full time education, six 17-year-olds in full time education were randomly selected from the control group and were included in the control group sample. A ratio of 1:6 was chosen as a balance of maintaining control group sample size and obtaining a well-matched sample with no statistically significant differences.

This process resulted in a control group sample of 1,043 participants with age and highest qualification held characteristics which matched those of the participants who engaged with the education course intervention. There were four cells where there was an insufficient number of control group participants in that cell to achieve the desired 1:6 ratio of engagers to control participants. This is why there were 1,043 control group participants rather than 1,086 (181 x 6).

Driver2020 - an evaluation of interventions designed to improve safety in the first year of driving

The Driver2020 study evaluated five interventions designed to reduce collisions and improve safety in young novice drivers in their first year of driving. The study was split into two arms; one covered interventions delivered to learner drivers, and the other covered interventions delivered to novice drivers (after they had passed their test). This report covers the learner arm. Three interventions were delivered to learner drivers who were aged 17-24 when they began learning. The interventions were a logbook app designed to increase on-road practice before test pass, a hazard perception training intervention designed to improve hazard perception skill, and an education intervention designed to provide learners with knowledge and skills, and ongoing monitoring strategies, to make them safer as drivers. Learner drivers who registered for the study were assigned randomly to one of the three intervention groups, or to a control group. The intervention participants were then invited to engage with the intervention if they wished. All participants completed self-report surveys after they passed their test, and at 3, 6 and 12 months post-test. These surveys asked people to report the driving they undertook, any collisions post-test, and various surrogate measures of safety targeted by the interventions. None of the interventions were associated with a reduction in collisions, and engagement levels with the interventions were very low. In participants who engaged, only the hazard perception training intervention was associated with clear safety improvements. Although the effects were small, this intervention was associated with a reduction in self-reported frequency of speeding in their first three months of driving, and a reduced number of attempts to pass the Driving Theory Test (which suggests a higher level of hazard perception skill).

Other titles from this subject area

PPR2009 Weekley J, Helman S, Chowdhury S, Hammond J and Hutton J (2024a). Driver2020 – an evaluation of interventions designed to improve safety in the first year of driving. Report D1: Effectiveness of interventions delivered to novice drivers. Commissioned and funded by the Department for Transport. Crowthorne: Transport Research Laboratory.

PPR2011 Hitchings J, Holcombe A, Christie N, Weekley J and Helman S (2024b). Driver2020 – an evaluation of interventions designed to improve safety in the first year of driving. Report D3: Delivery of interventions and engagement with them by novice and learner drivers. Commissioned and funded by the Department for Transport. Crowthorne: Transport Research Laboratory.

PPR2012 Helman S and Weekley J (2024). Driver2020 – an evaluation of interventions designed to improve safety in the first year of driving. Report D4: Summary of findings. Commissioned and funded by the Department for Transport. Crowthorne: Transport Research Laboratory.

PPR781 Pressley A, Fernández-Medina K, Helman S, McKenna FP, Stradling S and Husband P (2016). A review of interventions which seek to increase the safety of young and novice drivers. Commissioned and funded by the Department for Transport. Crowthorne: Transport Research Laboratory.

TRL
Crowthorne House, Nine Mile Ride,
Wokingham, Berkshire, RG40 3GA,
United Kingdom
T: +44 (0) 1344 773131
F: +44 (0) 1344 770356
E: enquiries@trl.co.uk
W: www.trl.co.uk

ISSN: 2514-9652

doi: 10.58446/sjtb9387

PPR2010