RAC Foundation

Mobility · Safety · Economy · Environment

DriveFit 2.0

Trial evaluation report

Dr Elizabeth Box
RAC Foundation

January 2026

The Royal Automobile Club Foundation for Motoring Ltd is a transport policy and research organisation which explores the economic, mobility, safety and environmental issues relating to roads and their users. The Foundation publishes independent and authoritative research with which it promotes informed debate and advocates policy in the interest of the responsible motorist.

RAC Foundation
89–91 Pall Mall
London
SW1Y 5HS

Tel no: 020 7747 3445
www.racfoundation.org

Registered Charity No. 1002705

RAC Foundation

Mobility · Safety · Economy · Environment

DriveFit 2.0

Trial evaluation report

Dr Elizabeth Box
RAC Foundation

January 2026

About this report

This report presents the evaluation of DriveFit 2.0, a redesigned pre-driver road safety intervention informed by the Pre-driver Theatre and Workshop Education Research (PdTWER) project. Building on the original DriveFit programme, DriveFit 2.0 was adapted for easier delivery in schools and colleges, combining active learning with film-based content. A trial evaluation, undertaken with Surrey Fire and Rescue Service, assessed its impact on young people's attitudes and intentions toward safe driving. The project, funded by the Road Safety Trust and supported by the RAC Foundation, National Fire Chiefs Council (NFCC), National Police Chiefs Council (NPCC), Devon and Somerset Fire and Rescue Service and Road Safety GB, ran from February 2024 to September 2025. The findings are intended to inform policymakers, road safety professionals, and educators about effective approaches to reducing young driver risk and strengthening pre-driver education in Great Britain.

About the author

Dr Elizabeth Box is an accomplished transport researcher and commissioner with over two decades of experience and an excellent track record in influencing and contributing to national road safety policy. She is the Research Director at the RAC Foundation, an independent charity researching the environmental, mobility, safety and economic issues relating to roads and their users. Elizabeth is also a behavioural science consultant at Co-Pilot, a start-up developing evidence-based road safety education, and Director of ECM Research Solutions, a specialist consultancy in behavioural science, intervention design, and evaluation.

Elizabeth holds a PhD in Transport Psychology from Cranfield University, where she developed and trialled pre-driver education interventions grounded in behavioural science. In 2025, she was honoured with Cranfield University's Faculty of Engineering and Applied Sciences Excellence in Achievement Alumni Award for Recent Safety Graduates, recognising the cross-disciplinary impact of her applied behavioural science research.

She advises several industry and government project boards, is a Fellow and technical champion of the Chartered Institution of Highways and Transportation, and serves as a judge for both the Prince Michael International Road Safety Awards and CIHT's annual Road Safety Award.

Disclaimer

This report has been prepared on behalf of the RAC Foundation by Dr Elizabeth Box. Any errors or omissions are the author's sole responsibility.

Contents

Figures list

Tables list

Executive summary

DriveFit 2.0 is a pre-driver road safety programme designed for delivery in schools and colleges. It builds on the original DriveFit intervention by combining selected film content with three structured, teacher-led lessons. The lessons focus on (1) driving practice and hazard prediction, (2) vehicle safety and managing distractions, and (3) managing fatigue and speed.

The programme is strengths-based, aiming to build skills, judgement, and practical strategies rather than relying on shock tactics. It is also designed to fit naturally within PSHE or enrichment provision, supporting scalable delivery without the need for external facilitators.

This evaluation, delivered in partnership with Surrey Fire and Rescue Service and funded by the Road Safety Trust, examined whether DriveFit 2.0 can be delivered by teachers in real school and college conditions, whether it is acceptable and engaging for students, and whether it shows any short-term changes in psychological outcomes linked to safer driving. These outcomes included attitudes and intentions relating to mobile phone use, fatigue, and speeding, as well as perceived risk and self-efficacy (i.e. students' confidence in their ability to manage risk and apply safe driving behaviours).

The study used a quasi-experimental pre–post design with a wait-list comparison group across four Surrey schools/colleges. Year 12 students (aged 16–17) completed surveys before and after delivery. The baseline sample was 442 students, with 145 students matched across baseline and follow-up (80 in the intervention group and 65 in the comparison group).

Overall, DriveFit 2.0 proved feasible to deliver as a teacher-led, curriculum-embedded programme. Delivery broadly followed the intended structure, based on observations and feedback from those supporting the programme. Students rated the lessons as credible and useful, and moderately engaging. Emotional responses were neutral to moderate rather than fear-inducing, suggesting the approach can prompt reflection while remaining aligned with PSHE best practice.

In terms of short-term outcomes, DriveFit 2.0 did not produce changes in attitudes, intentions, or self-efficacy once baseline levels and comparison-group trends were taken into account. Because outcomes were assessed only at a single follow-up point after the three lessons were completed, any immediate or short-lived effects occurring during or shortly after delivery may not have been detected. Students in both groups also started from relatively risk-averse positions, which reduced the scope for measurable short-term improvement. Several outcomes moved in a direction consistent with safer behaviour among intervention students. This pattern was observed within an intervention group that was predominantly male, but these shifts were modest and not statistically distinct from changes seen in the comparison group once initial scores were taken into account.

The only clear quantitative effect was a small but statistically significant increase in perceived risk. Intervention students rated risky driving scenarios as slightly more risky at follow-up than comparison students. While modest, this shift is consistent with DriveFit 2.0's aim of supporting more realistic risk awareness and judgement, and represents the clearest short-term signal of potential intervention impact in this evaluation.

Findings about delivery and acceptability were encouraging. Students responded similarly across genders, including young men who are typically at higher crash risk and sometimes less responsive to education-based programmes. Delivery generally followed the intended structure, based on observations and feedback from those supporting the programme. However, delivery was not checked in the same way across all schools, so it is not possible to confirm that the programme was delivered consistently in every setting.

Set alongside the original DriveFit cluster randomised controlled trial, which showed broader positive effects under a professionally facilitated model, DriveFit 2.0 appears to yield weaker short-term psychological change under teacher-led delivery. The present evaluation cannot isolate whether this difference is due to facilitation intensity, content, sample/context, or methodological factors such as power and follow-up timing. However, it highlights a key trade-off for policy and practice: teacher-led delivery supports scalability and sustainability, but may require additional supports or hybrid approaches if stronger impacts on behavioural determinants are sought.

Recommendations for refinement focus on strengthening how the programme helps students develop realistic awareness of driving risks and make sound judgements in risky situations, particularly through experiential and scenario-based learning activities. Clearer teacher support and simple monitoring tools (such as concise facilitation prompts, troubleshooting guidance, and short checklists or flowcharts) would help teachers deliver the lessons consistently as intended. Future evaluations should build on this initial work through longer-term designs, direct comparisons of delivery models (teacher-led, externally facilitated, and hybrid), and a wider range of outcomes, including behavioural or proxy indicators such as simulator tasks, telematics, or parent reports.

In conclusion, this evaluation provides a strong foundation for understanding DriveFit 2.0's feasibility, reception, and early effects in schools and colleges. It demonstrates that the programme can be embedded in routine provision, is acceptable to students, and is associated with a small positive shift in perceived risk. As the programme continues to develop, evaluation over longer timeframes and across different delivery models will be essential to strengthen the evidence base and clarify how DriveFit 2.0 and similar approaches can best support safer driving behaviours among young people.

1. Introduction

Road traffic injury remains a leading cause of death among 5–29-year-olds worldwide (WHO, 2023), with young and novice drivers in Great Britain disproportionately represented in crash statistics. Despite this persistent risk, traditional approaches to pre-driver education have often shown limited effectiveness, with many relying on fear-based appeals that are poorly aligned with best practice in behaviour change and classroom delivery. There is therefore a continued need for scalable, evidence-based interventions that can be feasibly delivered within schools and colleges, which engage young people in a constructive and sustainable way.

The DriveFit initiative was originally developed as part of the Pre-driver Theatre and Workshop Education Research (PdTWER) project, combining film-based content with professionally facilitated workshops to promote safer attitudes and intentions among pre-drivers. A cluster randomised controlled trial (cRCT) of the original programme demonstrated small but significant improvements in attitudes, intentions, and self-efficacy (i.e., confidence in one's ability to apply safe driving behaviours), alongside high levels of acceptability (Box & Dorn, 2023). These findings provided encouraging evidence of impact, but delivery at scale was constrained by the need for external facilitation.

DriveFit 2.0 was designed to address these challenges by adapting the intervention for teacher-led delivery embedded within the curriculum. The revised model retains the original behavioural science principles and key film content, but structures them into three classroom lessons supported by detailed teacher guidance, interactive activities, and student pledges. This evolution aims to make the programme more sustainable, scalable, and compatible with the Personal, Social, Health and Economic (PSHE) education context in schools.

The present evaluation, delivered in partnership with Surrey Fire and Rescue Service and funded by the Road Safety Trust, set out to examine the feasibility, acceptability, and short-term outcomes of DriveFit 2.0. Specifically, it sought to assess whether the intervention could be delivered effectively by teachers in schools and colleges, whether students found it credible and engaging, and whether it had measurable effects on psychological determinants of road safety such as attitudes, intentions, self-efficacy, and perceived risk.

The evaluation was deliberately designed as a smaller-scale, resource-constrained study, employing a quasi-experimental pre–post design with a wait-list control group. Its primary purpose was not to provide definitive tests of effectiveness but to generate insights into feasibility, acceptability, and immediate outcomes, while identifying lessons to inform programme refinement and future large-scale evaluation.

2. Method

This section outlines the methods used to design, deliver, and evaluate the DriveFit 2.0 intervention. It begins by describing the intervention itself, including its origins, theoretical foundations, and lesson content. The study design and experimental conditions are then detailed, followed by an overview of the measurement approach, survey content, process evaluation methods, and analytic procedures. Together, these subsections provide a transparent account of how the evaluation was conducted and how evidence on the effectiveness of DriveFit 2.0 was generated.

2.1 DriveFit 2.0 intervention

The DriveFit 2.0 intervention formed the core focus of this evaluation. The subsections below describe its development and structure, beginning with the origins of the original DriveFit programme and the rationale for revision. The updated approach, teacher guidance materials, and lesson content are then set out in detail, highlighting how DriveFit 2.0 adapts and extends the original intervention to ensure scalability and alignment with best practice in classroom delivery.

2.1.1 Origins of DriveFit 2.0

The original DriveFit intervention was developed in 2021 as a two-part programme for pre-drivers (aged 16-18 years old). It consisted of:

- A 40-minute film, delivered in the classroom, which introduced students to the risks and protective strategies relevant to young drivers.
- A 45-minute online facilitated workshop, delivered within two weeks of the film, led by professional facilitators using a structured framework to help students reflect on the film and apply the learning to their own lives.

The film was grounded in behavioural science and made use of Behaviour Change Techniques (BCTs) (Michie et al., 2013). It adopted a talk-show interview format, where expert guests provided demonstrations and advice on how young drivers could best manage the process of learning to drive and avoid risky behaviours such as speeding, fatigue, mobile phone use, and drink/drug driving. The content was informed by insights into the beliefs and experiences of the target audience, ensuring that the messages were relevant and relatable.

The workshop which followed extended the impact of the film by encouraging students to recall its content, review the key themes, and consider how they might apply the advice in practice. This was achieved through interactive polls, paired and group discussions, and scenario-based reflection. A central component of the workshop was the development of 'if–then' implementation intention plans, which supported students in anticipating barriers (e.g., peer pressure to speed) and identifying practical strategies to overcome them. Students were also invited to commit to their intentions by completing a DriveFit pledge card. Implementation intentions have been shown to help people translate good intentions into safer behaviours by providing a mental "if this happens, then I will do that" framework (Brewster et al., 2015; Gollwitzer, 1999). A supporting website provided additional resources for students, parents, and guardians[1].

2.1.2 The DriveFit 2.0 approach

DriveFit 2.0 builds on the foundation of the original DriveFit programme. The same core topics and messages are retained, with selected sections of the original film incorporated directly into the new materials. However, the delivery model has been substantially revised to ensure scalability, sustainability, and integration within schools. Instead of relying on external facilitators, DriveFit 2.0 provides:

- Three structured lesson plans (each 45–60 minutes), designed for use within PSHE or curriculum enrichment days.
- Teacher guidance notes with clear instructions, active learning strategies, and ready-to-use resources.
- A stronger emphasis on classroom-based interactive learning, including group discussions, quizzes, practical exercises, and student pledges.

1 The original trial website is no longer active. Full DriveFit programme details are available at https://roadsafetygb.org.uk/drivefit/.

This new model enables teachers to deliver DriveFit 2.0 independently, while maintaining fidelity to the evidence base and retaining the behavioural science principles that underpinned the original DriveFit intervention.

2.1.2.1 Teacher guidance

To ensure consistency and ease of delivery, DriveFit 2.0 includes a comprehensive Teacher Guidance pack. This provides:

- **Quick-start information:** Details about lesson length, class size (c.30 students), resources, and preparation steps.
- **Lesson format:** Each session includes learning intentions, mapped links to the PSHE Association Programme of Study, extension activities, take-away facts, and assessment grids.
- **Pedagogical approach:** Lessons are outlined as normative (data-led), strength-based, trauma-informed, restorative, and inclusive.
- **Best practice in PSHE:** The intervention supports establishing group agreements, creating a safe environment, using interactive methods, and providing additional signposted support.
- **Assessment tools:** Pre- and post-surveys, student self-assessments, teacher feedback grids, and reflective activities are used to capture impact.
- **Active learning toolkit:** Methods include diamond ranking, carousel discussions, quizzes, mind mapping, pledges, and solution-focused thinking are included in the lesson plans.

The teacher guidance ensures that teachers are supported to deliver lessons confidently, without requiring specialist road safety expertise, ensuring DriveFit 2.0 can be embedded at scale.

2.1.2.2 Lesson content

The DriveFit 2.0 intervention consists of three structured lessons, hosted on www.drivefit.info, each addressing a core aspect of road safety:

Lesson 1: Driving practice and hazard prediction

This session develops students' hazard perception skills and emphasises the importance of practice while learning to drive. Activities include:

- Skills for driving Diamond-16 ranking exercise
- Video-based hazard identification and group categorisation of risks
- Reflection on internal and external factors that influence hazard recognition
- Drawing and writing activities to capture perceptions of road risks

> **Take-away fact:** Hazard perception improves over time with practice and experience; becoming a safe driver takes time and practice, not just passing the test.

Lesson 2: Vehicle safety and managing distractions

This session focuses on making safe vehicle choices and managing common distractions. Activities include:

- Carousel activity on what makes a car safe
- Video analysis of crash test ratings, mobile phone use, and passenger influences
- Pledge writing around safer driving practices
- Group quizzes or a "five-star driver" exercise to consolidate learning

Take-away fact: Choosing the safest possible vehicle and planning in advance to manage distractions are essential to protecting both drivers and passengers.

Lesson 3: Managing fatigue and speed

This session addresses the risks of fatigue and speeding, encouraging students to prepare strategies in advance. Activities include:

- Sleep quiz to link rest with driving performance
- Video analysis on fatigue and safe driving
- "If this, then that" scenario planning for situations involving speeding pressures
- Solution-focused group discussions on overcoming barriers to safe behaviour
- Personal driving pledge and reflection activity

Take-away fact: Preparation and planning are crucial for resisting pressures to speed; safe driving requires judgement, reflection, and resilience.

Overall, DriveFit 2.0 retains the core evidence-based content and themes of the original intervention, while shifting to a teacher-led, classroom-embedded model that places active learning at the centre. By integrating film content into structured lesson plans, the intervention remains engaging while being more practical for schools and colleges to deliver. This evolution ensures that DriveFit continues to provide young people with the knowledge, skills, and attitudes needed for safer driving, while offering a sustainable and scalable model for road safety education nationwide. The teacher guidance note, the three lesson plans and supporting presentations can be accessed at www.drivefit.info.

2.1.2.3 Logic model for DriveFit 2.0

This logic model summarises the theoretical rationale underpinning DriveFit 2.0. Grounded in the Theory of Planned Behaviour (Ajzen, 1991) and informed by implementation intentions research (Gollwitzer, 1999), the programme targets key determinants of pre-driver behaviour, including attitudes, perceived behavioural control/self-efficacy and intentions. Through active learning tasks, facilitated reflection, risk appraisal, and planning exercises, DriveFit 2.0 is expected to strengthen these proximal determinants. These short-term psychological and capability-related changes represent the mechanisms through which safer behavioural intentions, and, in the longer term, safer novice driver behaviours, are expected to emerge.

Table 2-1: DriveFit 2.0 Logic Model

Inputs	Immediate impacts	Short-term impacts	Behavioural impacts	Health outcomes
Providing a three-lesson, teacher-led programme for 16–17-year-old pre-drivers, integrating film content, active learning tasks, and behavioural-science-based strategies, will…	Result in the delivery of DriveFit 2.0 within classroom PSHE/ enrichment sessions, with teachers using structured lesson plans and guidance, which will…	Strengthen students' understanding of key road risks (e.g., fatigue, speed, distraction), enhance perceived behavioural control/ self-efficacy (i.e., confidence in applying safe behaviours), support more realistic risk appraisal, and shape attitudes and intentions towards safer driving, which will…	Support the formation of safer behavioural intentions and the development of proactive strategies for managing common driver risks, ultimately contributing to…	Long-term reductions in risky driving behaviours among newly qualified drivers and, over time, reduced deaths and serious injuries among this high-risk group.

DriveFit 2.0 operationalises a set of Behaviour Change Techniques (BCTs) from the BCT Taxonomy v1 (Michie et al., 2013). Table 2-2 summarises the core BCTs used across the three lessons and illustrates how they are embedded within classroom activities. The list is not exhaustive, but highlights the techniques most central to the programme's theory of change.

Table 2-2: Behaviour Change Techniques (BCTs) incorporated in DriveFit 2.0

BCT (label)	BCT no. (BCTTv1)	How it is operationalised in DriveFit 2.0
1.0 – Goals and planning		
Problem solving	1.2	Solution-focused group discussions invite students to generate practical strategies for overcoming barriers to safe driving (e.g. resisting speeding pressures, managing passenger influence, improving habits) (Lesson 3).
Action planning (Including Implementation intentions)	1.4	The "If this, then that" worksheets guide students to create specific if–then plans for managing speeding pressures and other risky situations, translating general intentions into concrete responses (Lesson 3).
Commitment	1.9	Students are invited to write personal driving pledges, committing to specific safer behaviours (e.g. managing distractions, speed, and fatigue) and recording these in a form they can refer back to (Lessons 2 & 3).
2.0 – Feedback and monitoring		
Feedback on behaviour	2.2	Quizzes and group discussion tasks provide immediate feedback on students' understanding (e.g. plenary discussions across lessons).
Self-monitoring of behaviour	2.3	Self-assessment grids invite students to reflect on how far they have met the learning intentions, while the Sleep Quiz encourages reflection on personal sleep patterns and how these might affect driving (Lessons 1 & 3).

BCT (label)	BCT no. (BCTTv1)	How it is operationalised in DriveFit 2.0
4.0 – Shaping knowledge		
Instruction on how to perform the behaviour	4.1	Teachers and videos provide explicit guidance on safer driving-related skills, such as scanning for hazards, using mirrors, interpreting road signs, choosing safer vehicles, and managing distractions (Lessons 1–3).
Information about antecedents	4.2	Students discuss internal and external factors that influence hazard perception (e.g. fatigue, emotions, distractions), and explore how sleep patterns and daily variation in alertness affect driving performance (Lessons 1 & 3; Sleep Quiz).
5.0 – Natural consequences		
Information about health consequences	5.1	Video content and discussion highlight the consequences of speeding, fatigue, distraction, and unsafe vehicle choice (Lessons 2 & 3).
6.0 – Comparison of behaviour		
Demonstration of the behaviour	6.1	Video clips model safer practices, such as hazard prediction and distraction management, showing what "good" driving-related behaviours look like in context (Lessons 1–3).
Social comparison	6.2	Group tasks (e.g. Diamond-16 ranking of driving skills, carousel activity on what makes a car safe, shared quizzes) allow students to compare their own views and strategies with those of their peers (Lessons 1 & 2).
7.0 – Associations		
Prompts/cues	7.1	"Take-away facts", key messages on slides, and written pledges act as prompts and cues to support safer decision-making and recall of strategies beyond the lesson (all lessons).
8.0 – Repetition and substitution		
Behavioural practice/ rehearsal	8.1	Students mentally rehearse spotting hazards (video-based hazard identification), practise generating safer responses to risky situations, and rehearse implementation intention statements using the "If this, then that" activities (Lessons 1 & 3).
9.0 – Comparison of outcomes		
Credible source	9.1	Safety messages and guidance are delivered via expert voices and trusted sources (e.g. DriveFit video, EuroNCAP safety ratings, official government and RAC resources referenced in "Find out more" sections).

2.2 Study design and experimental conditions

The trial of DriveFit 2.0 was designed to assess the short-term effectiveness of the intervention in changing pre-driver attitudes and intentions. A quasi-experimental pre–post design with a comparison (wait-list control) group was employed, enabling outcomes to be compared between students who experienced the intervention and those who had not yet received it.

2.2.1 Participants

The evaluation focused on Year 12 students (aged 16–17) across four schools and colleges in Surrey. Recruitment was coordinated by Surrey Fire and Rescue Service, who approached schools they had existing engagement with and invited them to participate.

- **Intervention group:** Two schools/colleges participated, with a total of 197 students completing the baseline (T1) survey. Of these, 80 students completed the follow-up survey (T2).
- **Comparison (wait-list control) group:** Two schools/colleges served as the comparison group. At baseline, 245 students participated, with 65 completing the follow-up survey at the same timepoint as the intervention group. These students were provided access to the DriveFit 2.0 resources after the trial concluded.

Demographic information was collected from both groups, including age, gender, ethnicity, learning-to-drive status, driving practice intentions, hazard perception confidence, and attitudes towards safe driving behaviours. This allowed for a detailed profile of the samples and the ability to explore potential differences between groups.

2.2.1.1 Sample size rationale

No formal a priori power calculation was conducted for this evaluation. The achievable sample size was shaped by the number of schools and colleges willing to participate within the project timeframe and by the practical constraints of delivering three lessons during PSHE or enrichment sessions. As is typical in school- and college-based quasi-experimental evaluations, recruitment occurred at the class level rather than through individual sampling, resulting in group sizes that were naturally bounded by class enrolment. Previous evaluations of pre-driver and classroom-based behavioural interventions indicate that baseline samples of several hundred students generally provide a reasonable basis for exploring group differences and short-term changes in key psychological measures. Although attrition between baseline and follow-up reduced the matched analytic sample (80 intervention; 65 comparison), the available data were sufficient for examining short-term patterns of change, while recognising that smaller samples limit the ability to detect more subtle effects.

2.2.2 Delivery context

DriveFit 2.0 was delivered in classroom settings by teachers during timetabled PSHE or curriculum enrichment sessions. The intervention was entirely teacher-led, with only limited involvement from Fire and Rescue Service staff (e.g. briefing teachers in-person and observing a small number of sessions). All three lessons were requested to be delivered within a six-week period, though exact scheduling was determined by the participating schools/colleges.

2.2.3 Delivery model

The intervention comprised three sequential lessons (each 45–60 minutes), focusing on:

- Driving practice and hazard prediction
- Vehicle safety and managing distractions
- Managing fatigue and speed

All sessions were delivered by teachers, with delivery experience varying between schools/colleges (e.g. some delivered by PSHE staff, others by different subject teachers). To support delivery, teachers received:

- An in-person briefing from Surrey Fire and Rescue Service outlining the aims of the project.
- A facilitator guidance pack with detailed lesson plans, PowerPoint slides, and student activities.
- Letters of thanks to schools/colleges for their participation.

2.2.4 Timing

The trial was carried out during the Spring term (February – April 2025). Students in the intervention group completed a baseline survey immediately before Lesson 1 (T1) and a post-intervention survey after Lesson 3 (T2). The comparison group completed surveys at equivalent time points to allow for meaningful comparisons.

2.2.5 Experimental conditions

The evaluation compared outcomes between an intervention group (who received the three DriveFit 2.0 lessons) and a comparison (wait-list control) group who had not yet received the intervention. Students in the comparison group completed baseline and follow-up surveys at the same timepoints as the intervention group to allow meaningful comparison. After the trial period, DriveFit 2.0 resources were made available to these schools. While fidelity of delivery (i.e., how closely teachers followed the programme as intended) was not formally monitored across all sites, one session was observed by Surrey Fire and Rescue Service staff to provide insight into delivery. It is possible that lessons were not delivered entirely as intended in all settings, which should be borne in mind when interpreting the results.

2.2 Measurement approach

This section describes the methods used to collect, code, and analyse student and deliverer data for the DriveFit 2.0 evaluation. It outlines the survey administration procedures, the constructs measured, the approach to coding and scoring items, the process evaluation questions used with deliverers, and the analytic techniques applied to assess programme outcomes.

2.3.1 Survey administration

Student data were collected using online surveys hosted on SurveyMonkey. Baseline (T1) surveys were administered immediately before Lesson 1 in the intervention schools/colleges and at the same time in the comparison (wait-list control) schools/colleges. Post-intervention (T2) surveys were collected after Lesson 3 in the intervention schools/colleges and at the equivalent timepoint in the comparison schools/colleges. Although most surveys were completed in class time, follow-up with schools and colleges was required, with some students submitting their responses after the scheduled class session.

Surveys were accessed via QR codes displayed in lessons, with additional links distributed by Surrey Fire and Rescue Service coordinators through reminder emails. Pre–post responses were matched using students' first name, surname, and school/college name. Data were then anonymised following the matching process.

Informed consent was obtained from all participants prior to survey completion. Students were informed that participation was voluntary, responses would remain confidential, and they could withdraw at any point. Data handling followed GDPR principles, with identifiers removed once matching was complete to ensure anonymisation. Ethical considerations, including consent procedures and data protection, were reviewed and approved by the project Steering Board.

2.3.2 Survey measures

The survey instrument combined demographic items, measures of driving attitudes and intentions, indicators of perceived capability and risk awareness, and process evaluation questions. Table 2-3 provides an overview of the constructs measured, with the full survey included in Annex A.

Table 2-3: Summary of survey measures

Domain	Item details	Response format	Question numbers
Demographics & background	Gender, age, ethnicity; school/college attended; driving status; household vehicle access; anticipated hours of practice; expected frequency of practice; exposure to different driving environments	Multiple choice, categorical scales	Pre (Q1-10) Post (Q1-6)
Hazard-related measures	Confidence in ability to identify hazards while driving; actively looking for potential hazards when driving or riding as a passenger	7-point Likert scales (Not at all confident–Completely confident; Strongly disagree–Strongly agree)	Pre (Q11&12) Post (Q7 & 8)
Attitudes & intentions towards risky driving behaviours	Mobile phone use: attitudes and willingness/likelihood to engage; Fatigue: attitudes and willingness to drive when tired; Speeding: attitudes and willingness to drive over the limit	7-point scales using adjective pairs (e.g. harmful–beneficial) and Likert scales (Likely–Unlikely; Very willing–Not very willing)	Pre (Q13 – 18) Post (Q9 – 14)

Domain	Item details	Response format	Question numbers
Perceived safety of scenarios	Safety ratings for 10 scenarios (e.g. driving with two or more passengers, late-night driving, speeding, mobile phone use, driving older cars, drink/drug driving, going through a red light)	4-point scale (Always safe–Rarely safe)	Pre (Q19) Post (Q15)
Self-efficacy	Confidence in meeting the challenge of maintaining safe driving behaviours and managing personal risk whilst driving	5-point Likert scale (Strongly agree–Strongly disagree)	Pre (Q20) Post (Q16)
Process evaluation (Intervention group, post-survey)	Perceptions of DriveFit 2.0 lessons: credibility, usefulness, interest, importance, informativeness, enjoyment, surprise, shock, worry, fear; Insights and impact: new insights, personal benefit, relevance, behavioural influence; Additional comments.	5-point Likert scale (Strongly agree–Strongly disagree); open-text response	Post (Q17 – 19)

2.3.3 Survey item coding

Survey items were coded so that higher scores indicated more road-safety-supportive responses for outcome constructs (e.g., safer attitudes, greater confidence, stronger protective intentions). Process-evaluation constructs were scored in their natural direction (e.g., higher EMO = more negative emotion; higher face-validity/COG/ENG = stronger agreement). Several items required reverse coding to ensure consistent direction:

- Attitude adjective-pair items (e.g., Harmful–Beneficial) were reverse-coded so that higher scores reflected safer attitudes.
- The "boring" item in ENG was reverse-coded before forming the ENG composite.
- The EMO items (shocking, worrying, frightening) were not reverse-coded (higher indicates more negative emotional arousal).
- Perceived risk items were coded so that higher values represent greater perceived risk.
- The "active hazard scanning" item was not reverse-coded (higher indicates more active scanning).

Composite scores were created by averaging related items (attitudes, intentions, perceived risk, and process-evaluation scales). Internal consistency of composite scales was assessed using Cronbach's alpha. Open-text responses were retained separately for qualitative analysis.

2.3.4 Process evaluation questions for deliverers

In addition to student surveys, process evaluation data were also collected from those involved in supporting and observing the delivery of DriveFit 2.0 in schools and colleges. These questions were designed to capture insights into fidelity, reach, dose, recruitment, and contextual influences on delivery, in line with standard process evaluation frameworks. The aims were to understand how closely the intervention was delivered as intended, how engaged teachers and students were, and what barriers or facilitators shaped implementation. Table 2-4 provides an overview of the themes and sample questions posed to DriveFit 2.0 deliverers. The full question set is available in Annex B.

Table 2-4: Overview of process evaluation questions used with DriveFit 2.0 deliverers

Domain	Objective	Example questions
Fidelity: Implementation consistency	Understand how closely delivery matched the intended design	To what extent do you think the sessions were delivered as planned (structure, content, sequence)? Were any components skipped or adapted, and why? Did teachers feel adequately supported? How did student responses influence delivery? Overall delivery score (1–10).
Dose delivered: What was delivered	Determine completeness of intervention delivery	How many sessions were delivered in each school/college? Were all three completed in full? Were there challenges in ensuring full delivery? How was delivery tracked?
Dose received: Engagement and satisfaction	Assess engagement and perceived value by stakeholders	How engaged were students? Which aspects were most or least enjoyable or relevant? Were you satisfied with the quality and impact? What feedback was received from teachers or staff?
Reach: Who received the intervention	Understand how widely the intervention was accessed	How many students participated in each school/college? Were there patterns in attendance or engagement across groups?
Recruitment	Understand how schools and practitioners were recruited	How were schools/colleges selected? What worked well or created barriers in recruitment?
Context	Identify environmental and contextual factors influencing delivery	What factors helped delivery? What barriers or unexpected events hindered it? What recommendations would you give for future settings?

In addition, teacher feedback was sought via a short online survey covering the same process evaluation domains (fidelity, dose delivered/received, reach, recruitment, and context). However, uptake from teachers was very limited, with only one survey response received. While this limited the ability to draw systematic conclusions from teacher feedback, the response has been included qualitatively in the Results section as illustrative evidence.

2.3.5 Data preparation and analysis

Survey data were exported from SurveyMonkey into Microsoft Excel for initial formatting and preparation, and then imported into IBM SPSS Statistics for Windows, Version 30.0.0.0 for cleaning and analysis. Only complete surveys were included in the dataset. Pre–post responses were matched using students' name and school identifiers, after which the dataset was anonymised. Students who completed only the baseline (T1) or follow-up (T2) survey were excluded from longitudinal analyses but retained for cross-sectional descriptive analyses, which enabled comparisons between those who completed both timepoints (T1&T2) and those who only responded at T1. The final matched sample used for pre–post analyses comprised 80 students in the intervention group and 65 students in the comparison group, while the full baseline samples (197 intervention, 245 comparison) were used for descriptive comparisons. Because only complete surveys were retained, there were no missing items within the analysed dataset.

The analysis proceeded in four stages. First, descriptive statistics were calculated for demographic and background variables (see Table 3-1) to profile the intervention and comparison groups at baseline. Group differences were examined using chi-square tests for

categorical variables and two-way ANOVAs for continuous and scale measures. Second, a retention analysis explored whether demographic or psychological characteristics were associated with completing both T1 and T2 surveys, again using chi-square tests and two-way ANOVAs. Third, baseline comparability of psychological measures was examined using two-way ANOVAs across Theory of Planned Behaviour constructs (i.e., attitudes, intentions). Finally, primary outcome analyses focused on changes in attitudes, intentions, perceived risk, and self-efficacy between baseline and follow-up, comparing intervention and comparison groups. For these analyses, ANCOVAs were used, controlling for baseline scores, age, and gender, with effect sizes (partial η^2) calculated to estimate the magnitude of observed differences.

Process evaluation data were analysed descriptively, with Likert-scale items combined into composite scores and reported alongside qualitative feedback from open-text responses. Differences by gender in process evaluation outcomes were examined using ANOVAs. Process evaluation data from deliverers (Surrey Fire and Rescue staff insights and one teacher survey response) were treated qualitatively and summarised descriptively, recognising that uptake from teachers was very limited.

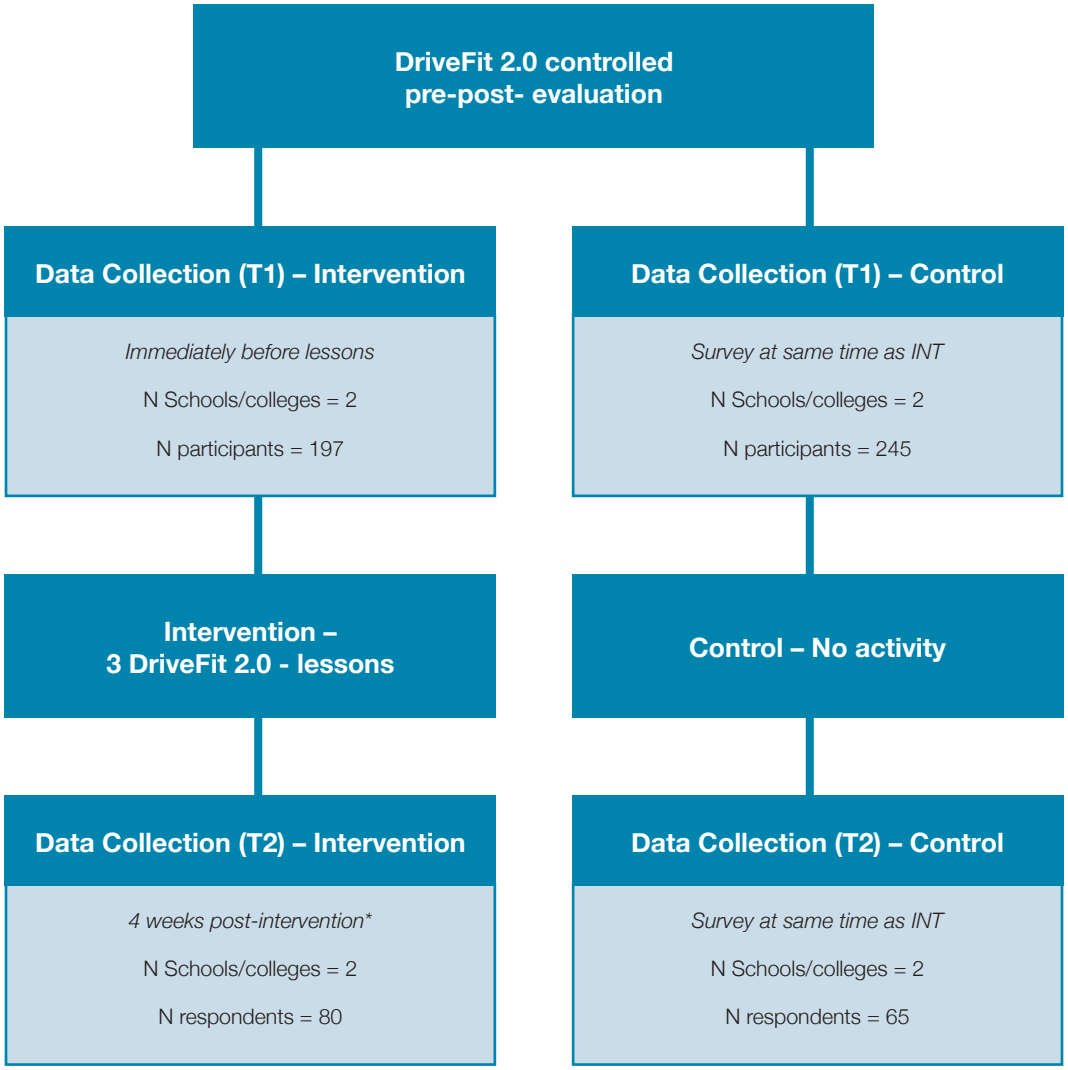
3. Results

This section reports the evaluation findings. Section 3.1 describes sample characteristics for the matched cohort (n = 145; 80 intervention, 65 control) and, where indicated, refers to full T1 cohorts. Section 3.2 presents outcome analyses comparing intervention and control groups at follow-up, adjusted for baseline, age, and gender. Section 3.3 summarises the process evaluation, including student survey results on acceptability and face validity, and qualitative insights from deliverers and a teacher respondent.

3.1 Sample characteristics and descriptives

Figure 3-1 illustrates the structure of the measurement sessions conducted for this evaluation, outlining the timeline and sequence of assessments. Table 3-1 outlines the descriptive statistics for the participants involved in the DriveFit 2.0 evaluation.

Figure 3-1: Measurement sessions for DriveFit 2.0 intervention evaluation



* Note. Time between T1 and T2 response (MEAN: 28 days, MAX: 72 days, MIN 7 days, MODE: 14 days)

Table 3-1: Descriptive statistics of participants in the DriveFit 2.0 evaluation (T1 & T2) (n = 145)

		Intervention group		Control group	
		T1 only	T1&T2	T1 only	T1&T2
N participants		117	80	180	65
Age (*SE*)		16.85 (.076)	16.80 (.096)	17.18 (.073)	17.12 (.121)
Gender	Male (%)	80 (68.4)	66 (82.5)	102 (56.7)	27 (41.5)
	Female (%)	36 (30.8)	9 (11.3)	66 (36.7)	37 (56.9)
	Non-binary or unknown (%)	1 (0.9)	5 (6.3)	12 (6.7)	1 (1.5)
Ethnicity	White (%)	70 (59.8)	49 (61.3)	126 (70.0)	44 (67.7)
	Non-white (%)	40 (34.2)	27 (33.8)	39 (21.7)	19 (29.2)
	Other/not stated	7 (6.0)	4 (5.0)	15 (8.3)	2 (3.1)
Learning to drive	Passed or currently learning	42 (35.9)	29 (36.3)	85 (47.2)	21 (32.3)
	Planning to learn (next 12 months – 5 yrs)	52 (44.4)	36 (45.0)	75 (41.7)	34 (52.3)
	Maybe or never learning to driver	23 (19.7)	15 (18.8)	20 (11.1)	10 (15.4)
Necessary driving hours before test (*SE*) (1 = 0-10hrs, 11 = 100+)1		3.94 (.225)	4.00 (.296)	4.12 (.197)	4.36 (.304)
Actual/planned driving practice per week (*SE*)(1 = Less than once a week, 5 = Everyday) 2		2.42 (.086)	2.28 (.089)	2.28 (.054)	2.42 (.085)
Exposure to different learning environments (*SE*) (1 = Never, 7 = Always) 3		4.23 (.130)	3.96 (.119)	4.17 (.108)	4.26 (.164)
Confidence in spotting hazards (as driver) (*SE*) (1 = Not at all confident, 7 = Completely confident)		4.88 (.134)	4.79 (.152)	5.07 (.109)	4.85 (.186)
Actively looking for hazards as driver/ passenger (*SE*) (1 = Strongly disagree, 7 = Strongly agree)		3.54 (.171)	4.35 (.212)	3.69 (.150)	3.66 (.252)
Confidence in ability to maintain safe driving behaviours (*SE*)(1 = Strongly disagree, 5 = Strongly agree)		3.72 (.130)	3.84 (.130)	4.02 (.092)	3.89 (.171)

1 Scale as follows – 1 = 0-10hrs, 2 = 11-20hrs, 3 = 21-30hrs, 4 = 31-40hrs, 5 = 41-50hrs, 6 = 51-60hrs, 7 = 61-70hrs, 8 = 71-80hrs, 9 = 81-90hrs, 10 = 91-100hrs, 11 = 100hrs +

2 Scale as follows – 1 = Less than once a week, 2 = 1-2 times a week, 3 = 3-4 times a week, 4 = 5-6 times a week, 5 = Everyday.

3 Scale as follows – 1 = Never, 2 = Very rarely, 3 = Rarely, 4 = Sometimes, 5 = Often, 6 = Very often, 7 = Always

Prior to examining intervention effects, analyses were conducted to test whether demographic characteristics were associated with follow-up response (T1-only vs. T1&T2). Gender was the only variable significantly related to retention. In the intervention group, a greater proportion of males (45.2%) and non-binary/unknown participants (83.3%)

completed both timepoints compared with females (20.0%; $\chi^2(2, N = 197) = 13.75$, $p = .001$). In the control group, the reverse pattern was observed: females were more likely to remain in the study (35.9%) than males (20.9%) or non-binary/unknown participants (7.7%; $\chi^2(2, N = 245) = 9.10$, $p = .011$). Ethnicity and learning-to-drive status showed no significant association with retention in either group (all ps > .20).

To extend this analysis, two-way ANOVAs were conducted to explore whether other baseline variables differed between those retained and those lost to follow-up. Three small but notable effects were identified:

- **Age:** Control group participants were slightly older (M = 17.15, SE = .067) than those in the intervention group (M = 16.82, SE = .067), $F(1, 438) = 12.29$, $p < .001$, partial $\eta^2 = .027$.
- **Hazard scanning:** There was a group x retention interaction, $F(1, 438) = 4.45$, $p = .035$, partial $\eta^2 = .010$, with higher scanning scores among intervention participants who completed both surveys compared with those who responded only at T1, and little difference in the control group.

No other measures showed significant differences by retention status. These findings suggest that while most baseline characteristics were comparable across response groups, some systematic differences existed, particularly in gender and certain driving-related measures, meaning that results should be interpreted with caution, as retention patterns may have introduced a degree of bias into the sample. Baseline comparability was also examined across Theory of Planned Behaviour (TPB) constructs (Table 3-2).

Table 3-2: Mean (SE) scores on Theory of Planned Behaviour constructs by study group (T1)[2]

	Intervention group		Control group	
	T1 Only	**T1&T2**	**T1 Only**	**T1&T2**
N participants	**117**	**80**	**180**	**65**
Mobile phone Attitudes (MOB_ATT)	5.89 (.115)	5.96 (.147)	5.93 (.109)	6.50 (.120)
Mobile phone Intentions (MOB_INT)	5.47 (.155)	5.74 (.181)	5.87 (.118)	6.15 (.157)
Fatigue Attitudes (FATIG_ATT)	5.90 (.104)	6.18 (.133)	6.05 (.086)	6.31 (.128)
Fatigue Intentions (FATIG_INT)	5.20 (.143)	5.62 (.166)	5.45 (.119)	5.73 (.176)
Speed Attitudes (SPEED_ATT)	5.79 (.122)	5.78 (.161)	5.77 (.113)	6.08 (.155)
Speed Intentions (SPEED_INT)	5.19 (.162)	5.43 (.187)	5.56 (.120)	5.74 (.192)

Note: Values represent mean (standard error) scores on a 1–7 scale, where higher scores indicate more road safety supportive responses.

2 Table 3-2 reports only those TPB constructs for which complete multi-item scales were available (attitudes and intentions). Although self-efficacy was measured, it represents only one component of Perceived Behavioural Control (PBC) and not the full PBC construct as conceptualised in the Theory of Planned Behaviour. For this reason, self-efficacy scores are reported separately in the broader outcome analysis rather than combined within the TPB table.

For mobile phone use while driving, students in the control group reported slightly safer attitudes than those in the intervention group ($F(1, 438) = 4.58$, $p = .033$, $\eta^2 = .010$). Students who went on to complete both surveys (T1&T2) also started out with slightly safer attitudes than those who completed only the first survey ($F(1, 438) = 5.57$, $p = .019$, $\eta^2 = .013$). There was no interaction between group and retention; in other words, the difference between the intervention and control groups was the same for students who stayed in the study and those who did not, meaning the two factors did not influence each other. Intentions followed the same pattern, with higher (safer) intention scores in the control group ($F(1, 438) = 6.25$, $p = .013$, $\eta^2 = .014$).

For fatigue-related attitudes, students who completed both surveys (T1&T2) reported slightly safer baseline attitudes than those who completed only the first survey ($F(1, 438) = 5.22$, $p = .023$, $\eta^2 = .012$). The same pattern was found for intentions, with safer baseline scores among students who remained in the study ($F(1, 438) = 4.90$, $p = .027$, $\eta^2 = .011$). However, unlike the mobile phone results, there were no differences between the intervention and control groups, and no interaction effects.

For speeding, there were no significant differences at baseline for either attitudes or intentions. Scores did not vary by group or by retention status, and no interactions were observed.

Taken together, these findings indicate that students who completed both T1 and T2 tended to report slightly safer attitudes and intentions at baseline than those who dropped out. This suggests that retention was not entirely random and should be considered when interpreting the intervention effects.

3.2 Outcome evaluation results

The outcome evaluation examined whether DriveFit 2.0 influenced students' expectations, attitudes, and self-reported behaviours relating to safe driving. Analyses compared intervention and control groups across baseline (T1) and follow-up (T2), focusing on practice-related measures (e.g., anticipated hours of learning and hazard perception), attitudes and intentions towards risky behaviours, and broader psychological constructs such as self-efficacy and perceived risk. Results are reported by outcome domain, with descriptive statistics and group × time analyses presented to identify any intervention effects.

3.2.1 Practice and hazard perception

This section explores whether DriveFit 2.0 influenced students' expectations about learning to drive and their confidence in hazard perception. Outcomes included anticipated hours of practice, expected frequency and variety of driving practice, confidence in identifying hazards, and active hazard scanning.

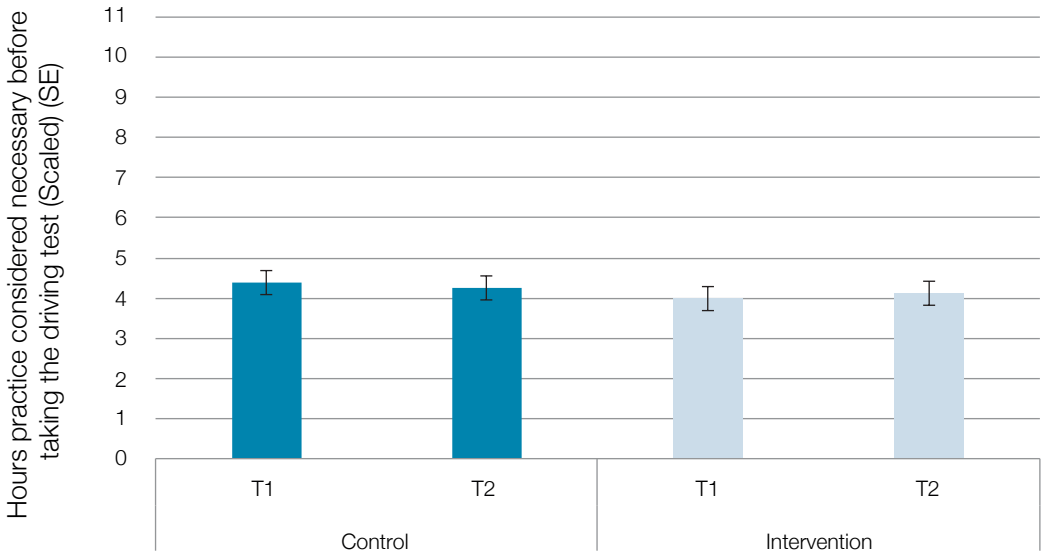
Across all measures, ANCOVA analyses (controlling for baseline, age, and gender) showed no significant intervention effects (all $ps > .24$). Instead, baseline responses consistently predicted follow-up scores, with moderate-to-strong effects (partial $\eta^2 = .15–.43$). Age was also associated with hazard identification confidence, with older students maintaining higher ratings. Table 3-3 summarises the ANCOVA results across outcomes.

Table 3-3: Summary of ANCOVA results for practice- and hazard-related outcomes at follow-up (T2)

Outcome	F(1,140) Group	p-value	partial η^2	Strongest predictor
Expected hours of practice	0.10	.748	.001	T1 expectations (η^2 = .429)
Frequency of weekly practice	0.59	.446	.004	T1 expectations (η^2 = .254)
Exposure to varied environments	1.36	.245	.010	T1 expectations (η^2 = .149)
Confidence in hazard identification	0.09	.763	.001	T1 confidence (η^2 = .115), Age (η^2 = .036)
Active hazard scanning	0.36	.551	.003	T1 scanning (η^2 =.028)

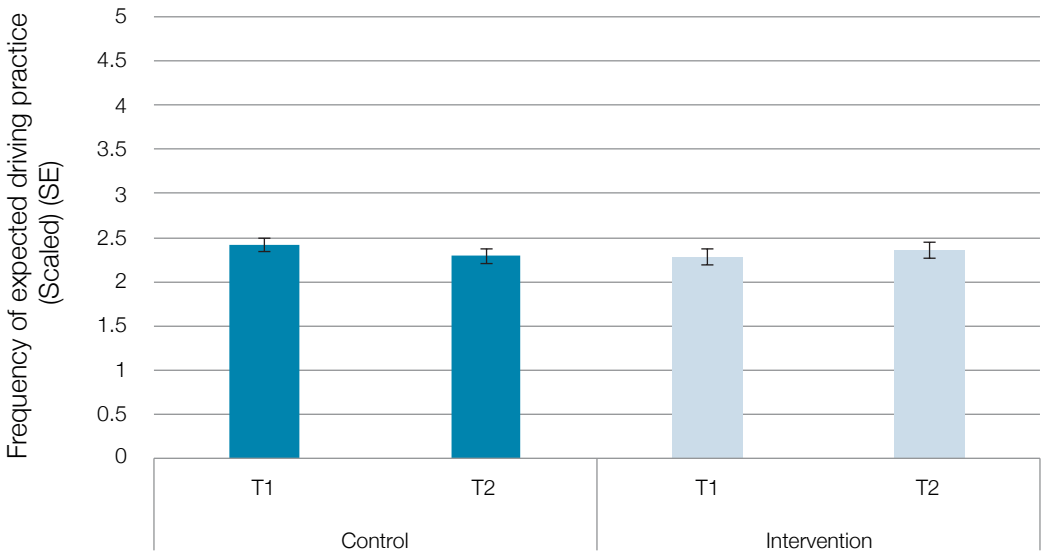
Figures 3-2 to 3-6 present the descriptive trends for each outcome. Both groups reported relatively stable expectations and confidence across timepoints, with no meaningful divergence. Full descriptive statistics are presented in Annex C.

Figure 3-2: Expected hours of driving practice before taking the test (T1 & T2), Intervention vs. Control*



* Scale as follows – 1 = 0-10hrs, 2 = 11-20hrs, 3 = 21-30hrs, 4 = 31-40hrs, 5 = 41-50hrs, 6 = 51-60hrs, 7 = 61-70hrs, 8 = 71-80hrs, 9 = 81-90hrs, 10 = 91-100hrs, 11 = 100hrs +
Note. Intervention group n = 80; Control group n = 65

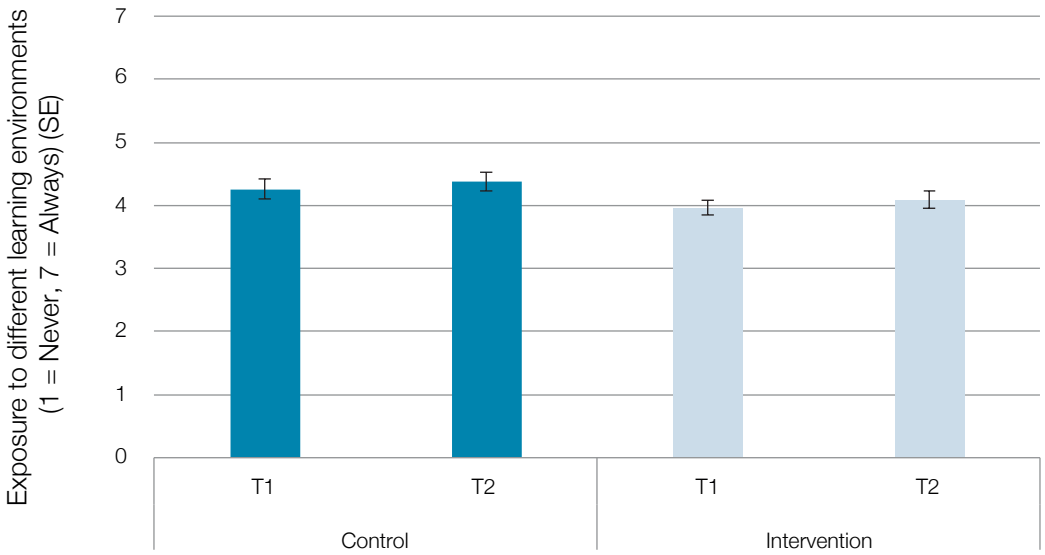
Figure 3-3: Expected weekly frequency of driving practice (T1 & T2), Intervention vs. Control*



* Scale as follows – 1 = Less than once a week, 2 = 1-2 times a week, 3 = 3-4 times a week, 4 = 5-6 times a week, 5 = Every day.

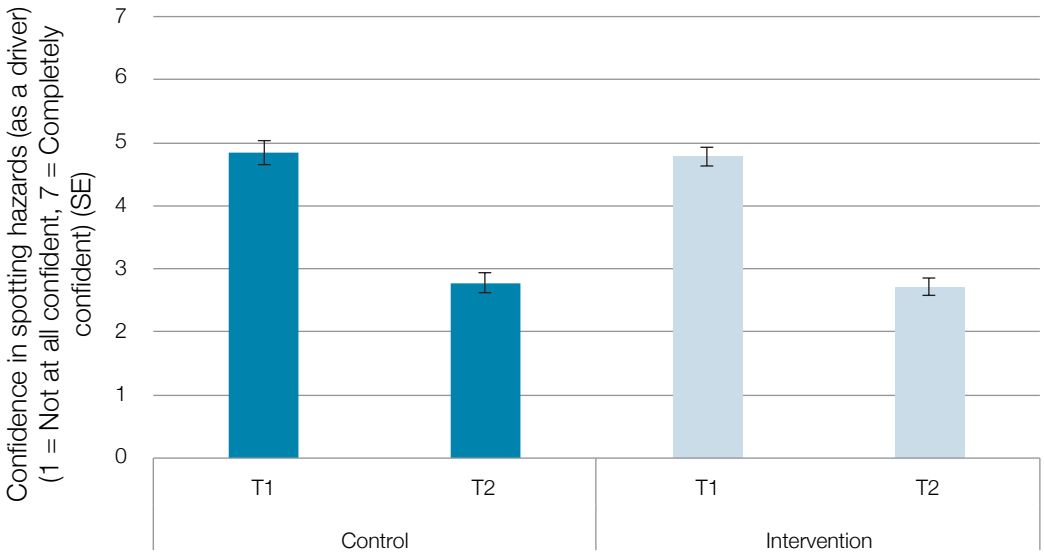
Note. Intervention group n = 80; Control group n = 65

Figure 3-4: Expected frequency of practising in different driving environments (T1 & T2), Intervention vs. control



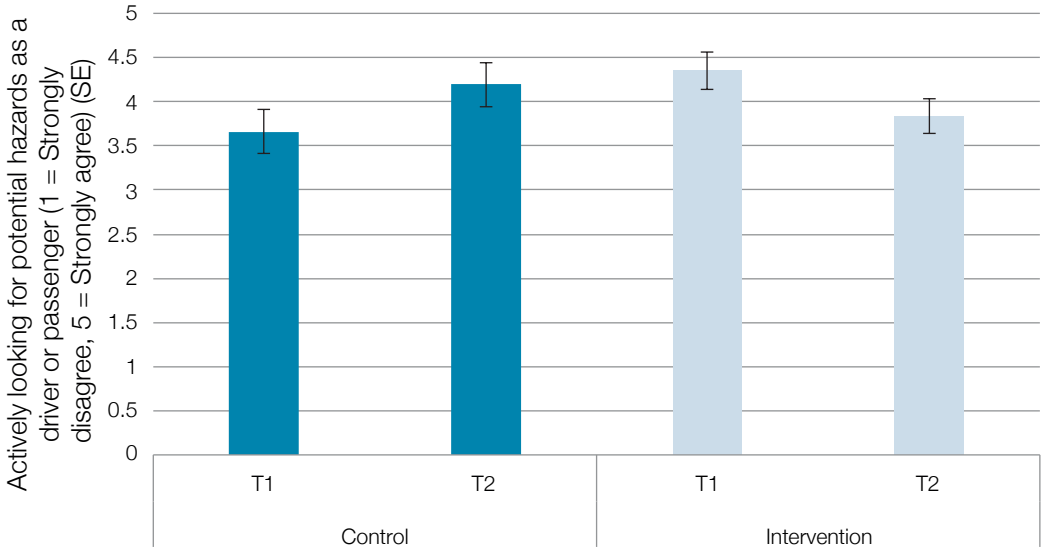
Note. Intervention group n = 80; Control group n = 65

Figure 3-5: Confidence in identifying potential driving hazards (T1 & T2), Intervention vs. control



Note. Intervention group n = 80; Control group n = 65

Figure 3-6: Active hazard scanning as a driver or passenger (T1 & T2), Intervention vs. control



Note. Intervention group n = 80; Control group n = 65

Although group effects were non-significant, some directional nuances emerged. Confidence in hazard identification declined across both groups, which may reflect a healthy calibration of self-assessment as students gain early learning experiences. Similarly, active hazard scanning decreased slightly among intervention students while increasing in controls,

possibly indicating that participation in DriveFit 2.0 heightened students' awareness of gaps in their own scanning behaviours, leading to more self-critical ratings.

Taken together, these results suggest that DriveFit 2.0 did not produce significant changes in practice-related expectations or hazard perception confidence. However, the small downward shifts seen in some confidence-related measures may indicate that students were reflecting more critically on their own developing skills. As this interpretation is observational rather than statistically confirmed, it should be treated as tentative.

3.2.2 Attitudes and behavioural intentions

This section examines whether DriveFit 2.0 influenced students' attitudes and intentions across three key risk domains: mobile phone use while driving, driving when fatigued, and speeding. Each construct was assessed with multi-item scales that demonstrated acceptable-to-good reliability across timepoints (see Table 3-4). Attitudes and intentions were analysed using ANCOVA, controlling for baseline scores, age, and gender.

Table 3-4: Internal consistency (Cronbach's alpha) for attitude and intention scales at baseline (T1) and follow-up (T2)

Domain	Attitudes (α)	Intentions (α)
Mobile phone use	T1 = .83, T2 = .73	T1 = .88, T2 = .82
Fatigue	T1 = .67, T2 = .82	T1 = .86, T2 = .85
Speeding	T1 = .80, T2 = .87	T1 = .89, T2 = .90

Across all three domains, no significant group differences were observed at follow-up (all $Fs < 1.1$, $ps > .29$). Instead, baseline scores were consistently the strongest predictors of T2 outcomes, with medium-to-large effect sizes (partial $\eta^2 = .13–.31$). Small demographic gradients were also apparent: males and slightly older students tended to report less risk-averse views, particularly for mobile phone use and fatigue.

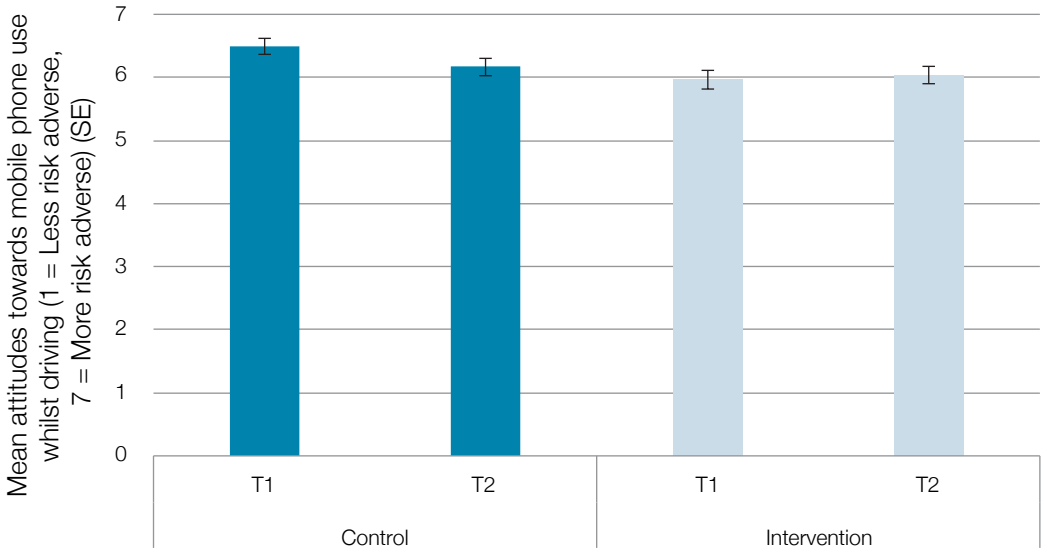
Table 3-5: Summary of ANCOVA results for attitudes and intentions at follow-up (T2), comparing interventions and control groups

Domain	Outcome	F(1,140) Group	p-value	partial η^2	Strongest predictor
Mobile phone	Attitudes	1.09	.299	.008	T1 attitudes ($\eta^2 = .128$)
	Intentions	0.07	.797	≈.000	T1 intentions ($\eta^2 = .250$)
Fatigue	Attitudes	0.47	.495	.003	T1 attitudes ($\eta^2 = .150$)
	Intentions	0.00	.973	.000	T1 intentions ($\eta^2 = .234$)
Speeding	Attitudes	0.39	.533	.003	T1 attitudes ($\eta^2 = .239$)
	Intentions	0.02	.892	.000	T1 intentions ($\eta^2 = .305$)

Although mean scores showed modest shifts between T1 and T2, these were not statistically attributable to the intervention. Figures 3-7 – 3-9 illustrate the descriptive trends

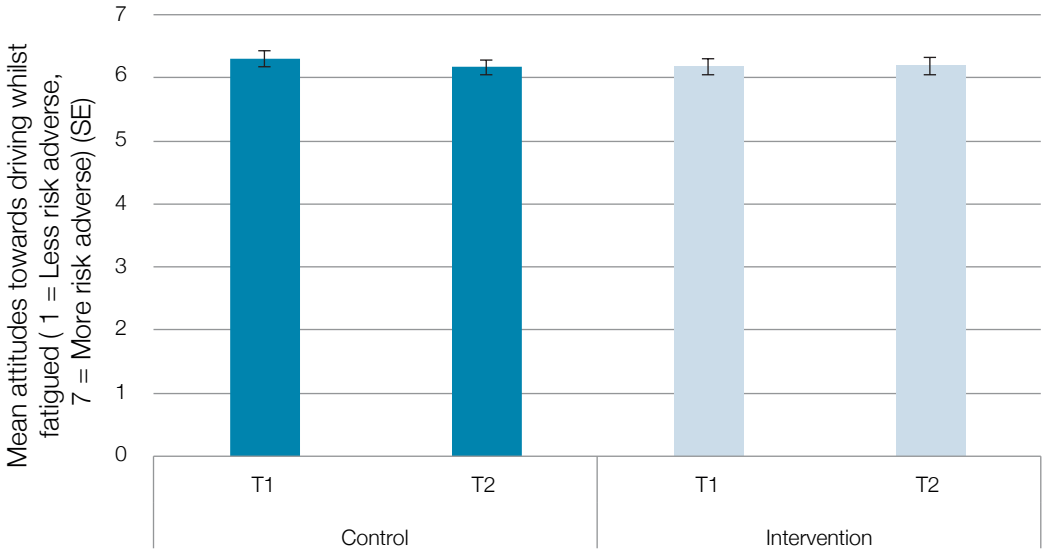
for attitudes, while Figures 3-10 – 3-12 show the corresponding trends for intentions. In each case, both groups reported relatively risk-averse views at baseline, with little meaningful divergence at follow-up. Full descriptive statistics are presented in Annex D.

Figure 3-7: Mean attitudes towards mobile phone use whilst driving at baseline (T1) and follow-up (T2), Intervention vs. Control.



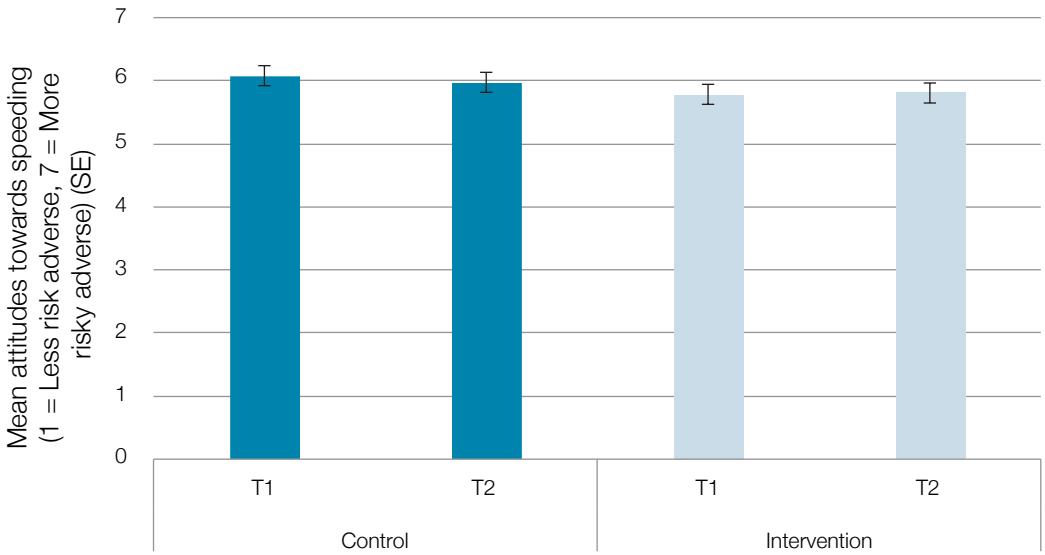
Note. Intervention group n = 80; Control group n = 65

Figure 3-8: Mean attitudes towards driving whilst fatigued at baseline (T1) and follow-up (T2), Intervention vs. Control.



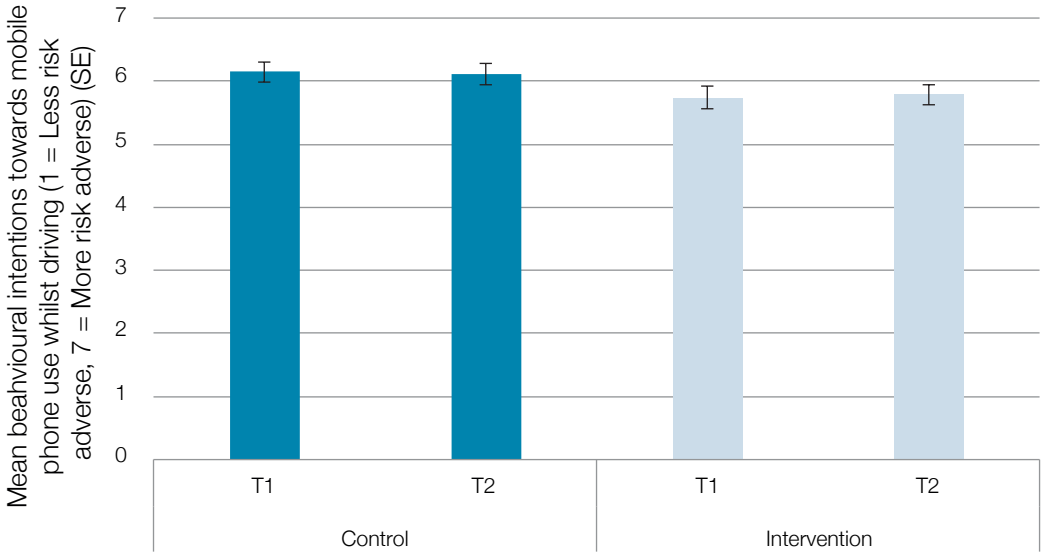
Note. Intervention group n = 80; Control group n = 65

Figure 3-9: Mean attitudes towards speeding at baseline (T1) and follow-up (T2), Intervention vs. Control.



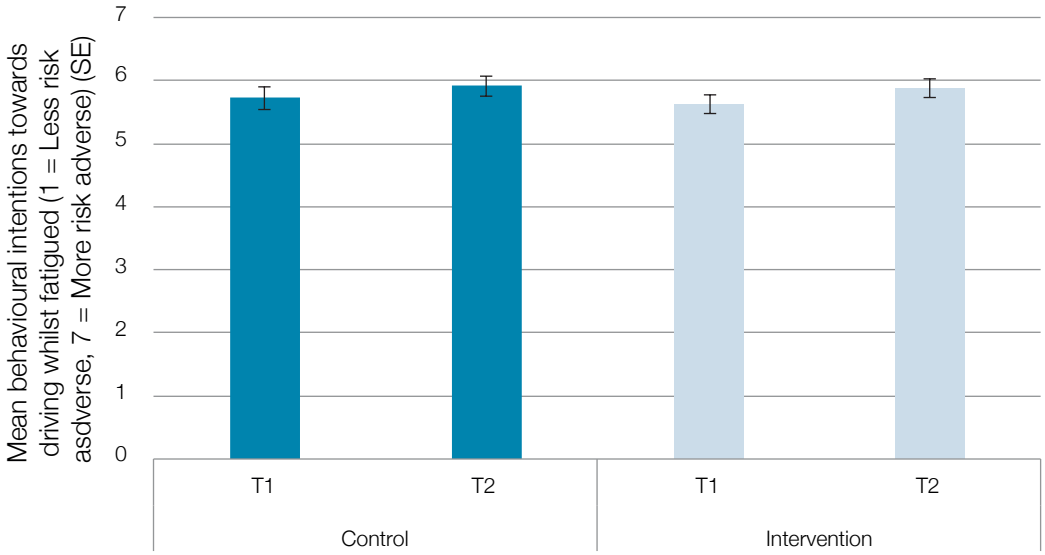
Note. Intervention group n = 80; Control group n = 65

Figure 3-10: Mean behaviour intentions towards mobile phone use at baseline (T1) and follow-up (T2), Intervention vs. Control.



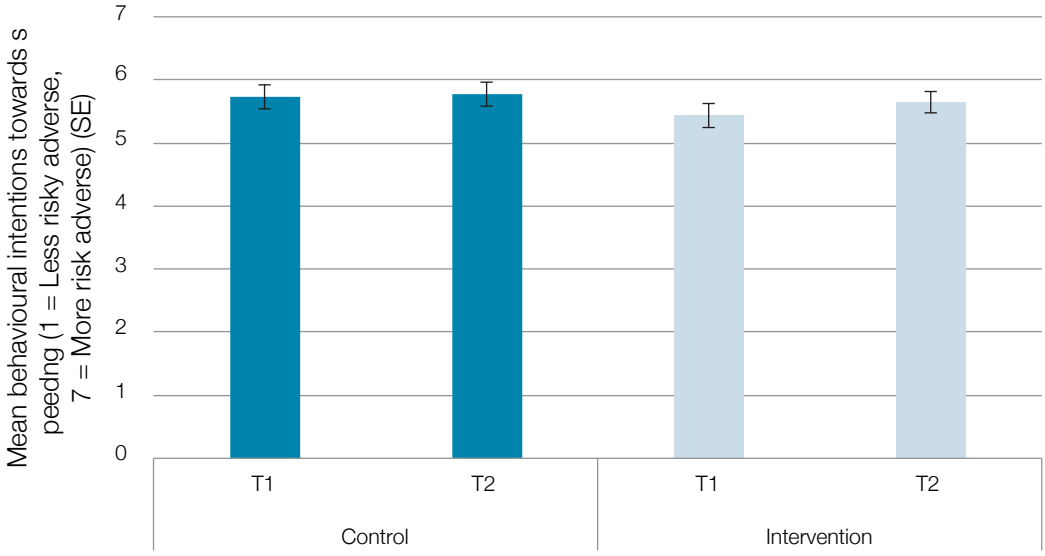
Note. Intervention group n = 80; Control group n = 65

Figure 3-11: Mean behaviour intentions towards driving whilst fatigued at baseline (T1) and follow-up (T2), Intervention vs. Control.



Note. Intervention group n = 80; Control group n = 65

Figure 3-12: Mean behavioural intentions towards speeding at baseline (T1) and follow-up (T2), Intervention vs. Control.



Note. Intervention group n = 80; Control group n = 65

Taken together, these findings suggest that DriveFit 2.0 did not significantly alter students' self-reported attitudes or behavioural intentions regarding mobile phone use, fatigue, or speeding. Students already held relatively risk-averse positions at baseline, limiting the potential for measurable intervention impact. Any changes observed over time were primarily

explained by students' initial responses, with only minor demographic influences. That said, descriptive trends (see Annex D) indicate that attitudes and intentions in the intervention group generally shifted in a more risk-averse direction than in the control group, where small declines or minimal change were observed. While not statistically significant, this pattern may suggest a more positive directional effect of DriveFit 2.0.

3.2.3 Self-efficacy and perceived risk

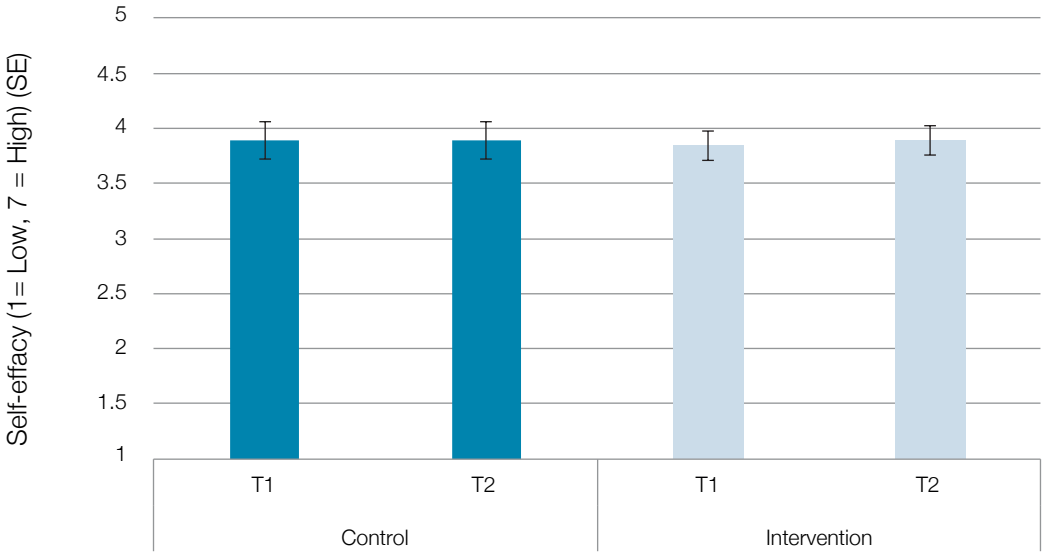
This section examines whether DriveFit 2.0 influenced students' self-efficacy for safe driving and perceived risk across common driving scenarios. Perceived risk was assessed with a 10-item scale that showed good internal consistency (overall $\alpha \approx .82–.85$). At T2, reliability remained strong in both groups (Control $\alpha = .85$; Intervention $\alpha = .84$). Self-efficacy was assessed with a single item. Perceived risk and self-efficacy outcomes were analysed with ANCOVA, and the results are summarised in Table 3-6.

Table 3-6: Summary of ANCOVA results at follow-up (T2): self-efficacy and perceived risk (Intervention vs. Control)

Outcome	F(1,140) Group	p-value	partial η^2	Strongest predictor
Self-efficacy	0.03	.868	≈.000	T1 self-efficacy ($\eta^2 = .069$)
Perceived risk	4.52	.035	.031	T1 risk perception ($\eta^2 = .351$)

Across both groups, self-efficacy remained relatively stable between baseline and follow-up, with very similar scores at T2 for the intervention ($M = 3.89$, $SE = 0.13$) and control groups ($M = 3.89$, $SE = 0.17$). ANCOVA results indicated no significant effect of group, $F(1,140) = 0.03$, $p = .868$, partial $\eta^2 \approx .000$, showing that DriveFit 2.0 did not differentially influence self-efficacy. Instead, baseline self-efficacy was the only meaningful predictor of follow-up scores, $F(1,140) = 10.45$, $p = .002$, partial $\eta^2 = .069$. Full descriptive statistics are presented in Annex E.

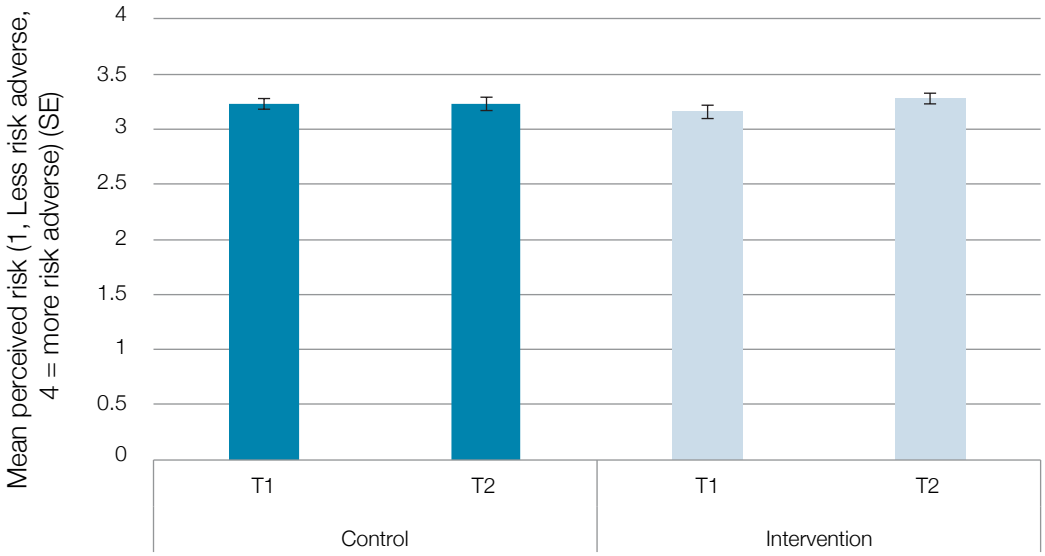
Figure 3-13: Self-efficacy at baseline (T1) and follow-up (T2), Intervention vs. Control.



Note. Intervention group n = 80; Control group n = 65

Perceived risk ratings were relatively high at baseline and remained stable in the control group (M = 3.23, SE = 0.06 at T2), but increased slightly in the intervention group (M = 3.28, SE = 0.05 at T2). ANCOVA revealed a small but statistically significant group effect, $F(1,140) = 4.52$, $p = .035$, partial $\eta^2 = .031$, with intervention students reporting greater perceived risk at follow-up compared to controls. Baseline perceptions were the strongest predictor of outcomes ($F = 46.99$, $p < .001$, partial $\eta^2 = .251$), and gender also played a role, with male students tending to rate risks as less risky overall ($F = 5.56$, $p = .020$, partial $\eta^2 = .038$).

Figure 3-14: Mean perceived risk at baseline (T1) and follow-up (T2), Intervention vs. Control.



Note. Intervention group n = 80; Control group n = 65

These findings show that DriveFit 2.0 did not influence students' self-efficacy, which remained stable over time and was almost entirely explained by students' baseline confidence levels. In contrast, perceived risk showed a small but statistically significant intervention effect, with DriveFit 2.0 participants rating risky driving scenarios as slightly more risky at follow-up than students in the control group. Although modest in size, this shift is consistent with the programme's aim of supporting more accurate risk appraisal. However, given the small effect and the strong influence of baseline perceptions, this result should be interpreted cautiously.

3.3 Process evaluation

The process evaluation focused on how DriveFit 2.0 was received by students and delivered in schools and colleges. This section reports on the acceptability of the programme, the consistency of delivery, and contextual factors that shaped implementation. Findings are presented in two parts: first, student survey feedback on perceptions of the intervention and its face validity, and second, insights from deliverer interviews and teacher feedback.

3.3.1 Student surveys

Student perceptions of DriveFit 2.0 were assessed through a post-session survey. The survey measured three aspects of programme acceptability: cognitive validity, engagement, and negative emotional responses, alongside a separate assessment of face validity (i.e., the extent to which the sessions appeared credible and relevant to participants). These measures provide insight into how well the intervention was received by its target audience and whether it was perceived as useful, engaging, and appropriate.

3.3.1.1 Perceptions of the intervention

Overall, student feedback on DriveFit 2.0 was moderately positive. Students generally agreed that the content was credible, useful, important, and informative, and engagement scores were slightly above the scale midpoint. Ratings of negative emotional arousal (e.g. finding the content shocking or worrying) were around the midpoint, indicating that the programme elicited some reflection without provoking strong fear-based responses. Across all scales, internal consistency was good to excellent (Cronbach's α = .84–.87). Unless otherwise noted, items were rated on a 5-point scale (1 = strongly disagree, 5 = strongly agree).

Table 3-7 summarises the descriptive statistics for each construct. Cognitive validity items received the highest ratings, with "important" (M = 4.11) and "informative" (M = 3.88) standing out. Engagement items (e.g. interesting, enjoyable) fell in the moderate range, and the reverse-coded "boring" item had the lowest endorsement (M = 2.80). Negative emotion items were rated between 3.14 and 3.38, reflecting neutral-to-moderate emotional responses.

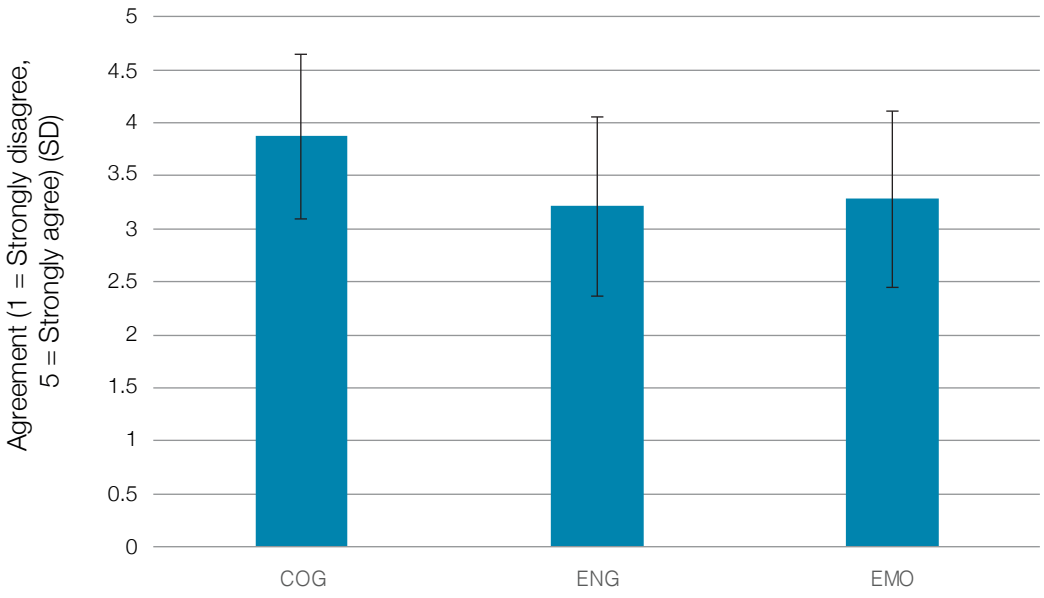
Table 3-7: Student survey results (item and scale level)

Construct	Items (M, SD)	Composite scale (M, SD, range)	α
Cognitive validity (COG)	Credible 3.72 (0.90); Useful 3.78 (0.96); Important 4.11 (0.81); Informative 3.88 (1.01)	3.87 (0.78, 1.25–5.00)	.86
Engagement (ENG)	Interesting 3.51 (1.04); Enjoyable 3.39 (1.13); Surprising 3.14 (1.00); Boring* 2.80 (0.97)	3.21 (0.85, 1.00–4.75)	.84
Negative emotion (EMO)	Shocking 3.14 (0.92); Worrying 3.30 (0.94); Frightening 3.38 (0.94)	3.28 (0.83, 1.00–5.00)	.87

Note: Items were rated on a 5-point scale (1 = strongly disagree, 5 = strongly agree). The "boring" item was reverse-coded prior to inclusion in the composite ENG scale.

Figure 3-15 provides a visual summary of the item-level means and standard deviations across the three constructs, illustrating the relatively higher ratings for cognitive validity, followed by moderate engagement scores and mid-range negative emotion ratings.

Figure 3-15: Mean scores (±SD) for cognitive validity (COG), engagement (ENG), and negative emotion (EMO) composite scales. Scores were rated on a 5-point scale (1 = strongly disagree, 5 = strongly agree).



Note. Intervention group n = 76

ANOVAs showed no significant gender differences for cognitive validity ($F(2,73) = 0.21$, $p = .812$) or engagement ($F(2,73) = 0.84$, $p = .436$). For negative emotion, the omnibus test approached significance ($F(2,73) = 3.06$, $p = .053$). Exploratory post hoc comparisons suggested that non-binary/unknown students reported higher emotional reactions than males or females (ps < .001), however, the very small size of this subgroup means this pattern should be interpreted cautiously and cannot be considered robust. No difference was observed between males and females ($p = .959$).

Overall, ratings of cognitive validity and engagement were broadly similar across gender categories within the intervention group. However, the sample was predominantly male, with limited representation of female and non-binary/unknown students, reducing the sensitivity of the analyses to detect gender-related differences. As such, the absence of statistically significant differences should not be interpreted as evidence of equivalence across genders.

Previous evaluations of road safety and health education interventions have sometimes reported higher acceptability among female students (e.g., Box & Dorn, 2023). Future evaluations with more balanced gender representation would be required to determine whether such patterns are also observed for DriveFit 2.0. At present, the findings suggest that DriveFit 2.0 was perceived as acceptable across the genders represented in the sample, without clear evidence of differential engagement.

A small number of students (n = 12) provided open-text comments. Most had no additional feedback; two offered brief remarks such as “Very good!” and “I can’t drive yet.”

3.3.1.2 Face validity

Face validity[3] was also assessed through a set of five items rated on a 5-point scale (1 = strongly disagree, 5 = strongly agree). The items asked whether the sessions provided new insights, were personally beneficial, covered relevant ideas, highlighted changes needed, and could make a difference to decisions and behaviours.

Mean ratings for these items were all slightly above the scale midpoint, ranging from 3.46 to 3.61, with a composite average of M = 3.53 (SD = 0.75, N = 76). The scale demonstrated excellent internal consistency (Cronbach’s α = .88), indicating that the items worked well together to assess perceived credibility.

Table 3-8 presents the item-level results. The relatively consistent scores across items indicate that students generally viewed the sessions as moderately relevant and useful, though not strongly so. Responses suggest that the programme was seen as offering some new insights and personal benefit, but with scope for stronger impact.

Table 3-8: Face validity item statistics

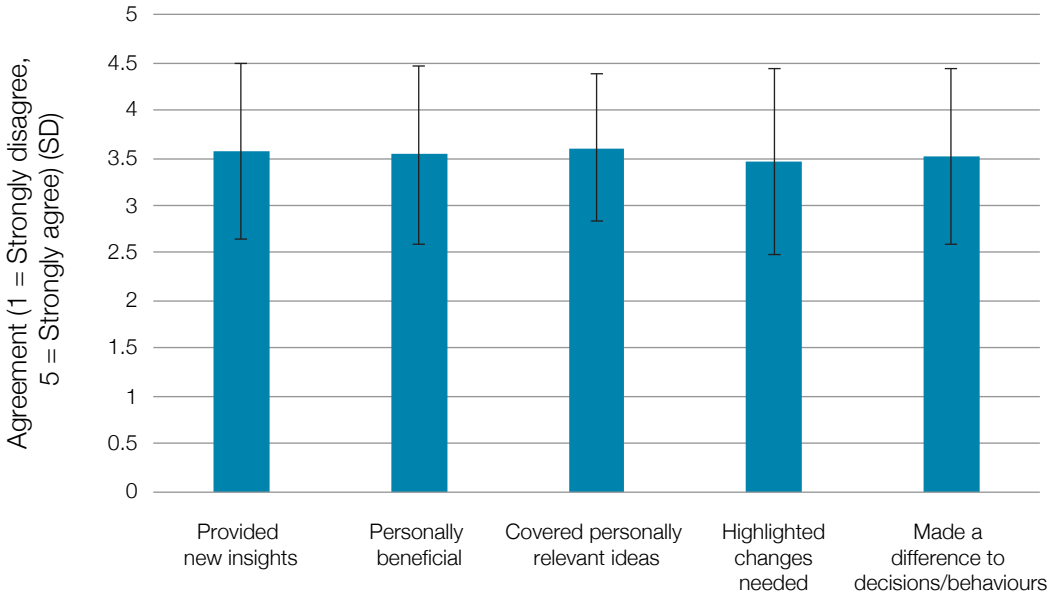
Item	Mean (SD)
Provided new insights	3.57 (0.91)
Personally beneficial	3.53 (0.94)
Covered personally relevant ideas	3.61 (0.78)
Highlighted changes needed	3.46 (0.97)
Made a difference to decisions/behaviours	3.51 (0.92)

Note. Items were rated on a 5-point scale (1 = strongly disagree, 5 = strongly agree). Higher scores indicate stronger agreement that the sessions were relevant, beneficial, and provided useful insights.

3 Face validity refers to the extent to which a measure or intervention appears, on the surface, to be credible, relevant, and appropriate to its intended purpose, as judged by participants.

Figure 3-16 summarises mean ratings across the five items, showing a consistent pattern of mid-to-upper-midpoint scores. This suggests that students felt the sessions offered some useful information and relevance, though endorsement was moderate rather than strong.

Figure 3-16. Mean ratings (±SD) for face validity items. All items were scored on a 5-point scale (1 = strongly disagree, 5 = strongly agree).



Note. Intervention group n = 80; Control group n = 65

Gender analyses of the overall face validity index (FCVALD) showed no significant differences ($F(2,73) = 0.004$, $p = .996$, $\eta^2 = .000$). Male, female, and non-binary/unknown students rated the intervention similarly, indicating broadly consistent perceptions of its relevance and usefulness across groups. However, the gender distribution within the intervention group was highly uneven, with relatively small numbers of female and non-binary/unknown participants, which limits the sensitivity of the analysis to detect potential gender-related differences.

3.3.2 Deliverer interviews

In addition to student survey data, qualitative insights were gathered from Surrey Fire and Rescue Service staff who supported or observed the delivery of DriveFit 2.0 in participating schools/colleges. These conversations explored fidelity, dose delivered and received, recruitment processes, and contextual factors influencing implementation. Teacher feedback was also sought through a short survey. Together, these perspectives provide valuable context for understanding how DriveFit 2.0 was delivered in practice and the factors that supported or hindered its implementation.

3.3.2.1 Fidelity and delivery consistency

Deliverers generally reported that schools followed the intended three-session structure of DriveFit 2.0. Teachers appeared to deliver the content in sequence, and there was no

evidence that sessions were routinely skipped. However, there were small adaptations to suit particular groups. For example, one teacher replaced a small-group exercise on car safety features with an individual "carousel" activity where students wrote ideas on paper and passed them around as "she felt it worked better with the group of boys she was working with".

These adjustments reflected the need for teachers to adapt to classroom dynamics rather than a disregard for the lesson plan. Deliverers emphasised that this balance between fidelity and flexibility is central to PSHE-style education, where "every class and every teacher is different" and absolute consistency is "really hard" to achieve.

Observed sessions were described as broadly faithful to the planned design, though differences in teacher experience shaped how smoothly lessons ran. Less experienced staff were sometimes less confident in facilitation, while those with more years in post appeared able to get more out of their groups. One deliverer also noted that trainee teachers were involved in delivery at one of the intervention schools, which may have introduced some variation in delivery quality or consistency, given differences in teaching experience. Overall fidelity was rated between 6 and 7 out of 10, with minor adaptations and some issues such as technology failures (e.g., films not playing) reducing the score. However, it should be noted that there was no formal mechanism in place to verify delivery fidelity across all schools. Observations were limited to one delivery session, and beyond this, fidelity was inferred from deliverer reflections and discussions with schools/colleges rather than being systematically assessed.

3.3.2.2 Dose delivered and dose received

All schools intended to deliver the full set of three sessions. Observations and follow-up conversations confirmed that lessons were timetabled week by week. However, it was not possible to verify exactly how many classes or individual students took part. The evaluation surveys suggested a degree of mismatch between pre- and post-participants, raising uncertainty about whether all respondents had completed the full programme.

Teachers reported that sessions were manageable to deliver and that materials were written in a familiar lesson-plan format, which helped reduce barriers. Feedback was generally positive: "It's been easy enough to follow," one deliverer reported hearing from multiple teachers. Students were seen to engage with tasks, with quieter individuals still contributing in writing, even if they were less vocal during discussion sessions.

3.3.2.3 Reach and recruitment

Recruitment of schools was achieved partly through existing networks (e.g., Healthy Schools) and partly through direct outreach. Deliverers felt that schools involved were proactive and engaged, which gave the intervention the best chance of being delivered as intended. Cohorts varied, however, with differences in educational level (e.g., FE college vs. sixth form) and gender balance between sites. This variation meant that delivery contexts were not uniform. One of the deliverers observed that "the schools are very different in their cohorts," and this influenced group dynamics.

No major barriers were reported in recruitment, though natural disruptions such as Ofsted inspections occasionally affected scheduling. Importantly, having supportive leads within schools made a significant difference, with one described as “on board from the start and happy to help us”.

3.3.2.4 Contextual influences

Context played a notable role in delivery. Technology was the most common barrier, with one school unable to show the intervention films due to sound issues with classroom computers. Facilities and classroom layouts also shaped delivery: in one case, computer-desks limited the ability to run small-group discussions.

At the same time, school culture and leadership support acted as key facilitators. Teachers who had engaged in prior training on moving away from fear-based methods were more invested in DriveFit 2.0’s strength-based approach. As one of the deliverers said, “he was on that same journey [as fire services in] understanding the move away from shock tactics”.

Broader contextual factors such as school culture, inspection cycles, and the timing of sessions within the school day also influenced delivery. Students were generally engaged, but behaviour and energy levels varied by class and teacher. Deliverers highlighted that such variability is typical of PSHE provision, where teaching is shaped more by teacher style and context than in more standardised subjects.

3.3.2.5 Teacher feedback

Only one teacher completed the feedback survey, which limits the ability to generalise findings. However, the response provides useful insight into how DriveFit 2.0 was perceived in practice.

The teacher rated the set-up and delivery of the lessons as “easy”, and agreed that the information provided beforehand was clear and straightforward to follow. They also agreed that students were able to engage with the interactive activities and film clips without difficulty.

On the question of content, the teacher indicated that lessons were interesting, though other items (credible, useful, informative, fun, engaging, boring) were left blank. When asked how well the lessons held students’ attention (1–10 scale), the teacher gave a score of 6, suggesting moderate but not universal engagement. Similarly, the content was judged to be pitched at about the right level, with a rating of 6 out of 7.

The teacher agreed that students benefited from the lessons and that the programme raised awareness of responsibilities both as drivers and as passengers. They also agreed that they would recommend DriveFit to colleagues and would support sharing information with parents and guardians.

No written comments were provided on areas for improvement, accessibility, or inclusivity. Nonetheless, this feedback indicates that the lessons were viewed as practical, moderately engaging, and beneficial, while also highlighting the need for stronger evidence from a larger pool of teacher respondents in future evaluations.

3.3.2.6 Summary

Overall, the deliverer interviews and teacher feedback suggest that DriveFit 2.0 was delivered broadly as intended, with schools making only minor adaptations to suit classroom dynamics. Teachers generally followed the planned three-session structure, and the materials were described as clear, manageable, and familiar in format. Students were observed to engage with activities, although levels of participation varied across classes and teacher styles. Recruitment benefited from existing networks and supportive school leads, while contextual factors such as technology issues, classroom layout, and inspection cycles occasionally disrupted delivery.

Fidelity was judged to be moderate-to-good, with ratings of 6–7 out of 10, but this conclusion rests largely on self-report and limited observation. Importantly, the evaluation did not include a formal mechanism for systematically monitoring fidelity across all sites. As a result, while delivery appeared consistent in most cases, it is not possible to confirm that the programme was implemented exactly as intended in every setting.

These findings highlight that DriveFit 2.0 can be delivered successfully in schools and colleges, with generally positive feedback from both staff and students. At the same time, they point to the importance of embedding stronger fidelity checks in future evaluations to provide greater confidence in how consistently the intervention is implemented.

4. Discussion

4.1 Summary of key findings

Across four Surrey schools/colleges, DriveFit 2.0 proved feasible to deliver as a teacher-led, classroom-embedded programme. Student perceptions were moderately positive: cognitive validity and engagement ratings were above the scale midpoint, negative emotional arousal was neutral-to-moderate, and face validity items were rated slightly above the midpoint. This indicates that students generally viewed the sessions as credible, useful and relevant, without experiencing strong fear-based reactions.

Outcome analyses found no differential effects on attitudes or intentions concerning mobile phone use, fatigue, or speeding once baseline levels were controlled. Students in both groups started from relatively risk-averse positions, and baseline attitudes and intentions were the strongest predictors of follow-up scores. Self-efficacy remained relatively stable between T1 and T2 and did not differ by group with follow-up confidence being largely explained by students' initial self-efficacy.

A small but statistically significant effect was observed for perceived risk at follow-up, with intervention students rating risky scenarios as slightly more risky than controls (partial $\eta^2 \approx .03$). Although modest, this effect is directionally consistent with DriveFit 2.0's aim of supporting more accurate risk appraisal. Given the small effect size and the strong influence of baseline perceptions, this finding should, however, be interpreted cautiously.

It is also noteworthy that the intervention group was predominantly male, a group that typically reports lower risk perception and less risk-averse views. Despite this, descriptive trends across several outcomes, including perceived risk, tended to move in a more safety-supportive direction in the intervention group than in the control group. These patterns were not statistically significant once baseline levels were controlled and should therefore be interpreted cautiously, but they indicate that DriveFit 2.0 was associated with favourable directional shifts within a sample that included a high proportion of students traditionally considered at elevated crash risk.

Analyses of practice-related expectations (anticipated hours and frequency of practice, exposure to different environments) and hazard-related measures (confidence in hazard identification and active hazard scanning) showed no significant intervention effects. Descriptively, hazard identification confidence declined slightly in both groups, and active hazard scanning fell marginally in the intervention group while increasing in controls. One interpretation is that participation in DriveFit 2.0 may have encouraged more critical self-assessment among intervention students, but as this interpretation is observational rather than statistically confirmed, it remains tentative.

Process findings further indicate that DriveFit 2.0 was generally implemented in line with its intended three-lesson structure, with teachers making only minor adaptations to suit classroom dynamics. Fidelity was judged to be moderate-to-good based on limited observation and deliverer reports, though the absence of systematic fidelity monitoring means implementation cannot be verified across all sites. Student engagement was described as reasonable overall but variable by class and teacher.

Participants reported similar levels of cognitive validity, engagement, and face validity across gender categories. However, the intervention group was predominantly male, with relatively few female and non-binary/unknown participants, meaning that these findings are descriptive rather than indicative of true gender differences. A tentative pattern of higher negative emotional responses was observed among the small non-binary/unknown subgroup, but this should be interpreted cautiously given the very small numbers involved.

These findings suggest that DriveFit 2.0 is feasible and acceptable as a teacher-led programme, with some promising directional effects on perceived risk but no clear short-term impact on attitudes, intentions, or self-efficacy in this small, relatively risk-averse sample. Retention patterns and the lack of comprehensive fidelity monitoring highlight the need for cautious interpretation and point to priorities for future, larger-scale evaluations.

4.2 Comparison with the original DriveFit trial

The findings from this evaluation can be considered alongside those of the original DriveFit programme, which was tested in a cluster randomised controlled trial (Box & Dorn, 2023). That more rigorous design provided stronger internal validity and demonstrated small but statistically significant improvements in road-safety-supportive attitudes, intentions, and self-efficacy, particularly for mobile phone use and speeding, alongside high acceptability ratings. The original model combined a classroom film with an externally facilitated workshop led by trained professionals, using structured discussion, reflection, and implementation intention planning.

In contrast, the present quasi-experimental evaluation of DriveFit 2.0, which shifted to a teacher-led, classroom-embedded delivery model, found no significant short-term effects on attitudes, intentions or self-efficacy once baseline levels were controlled. The only clear quantitative effect was a small increase in perceived risk, with intervention students rating risky scenarios as slightly more dangerous at follow-up than their control counterparts. Acceptability ratings were moderately positive in both trials, suggesting that each model was generally well received, but the detectable psychological impacts appear weaker under the revised teacher-led delivery.

Several factors may contribute to these differences. First, the designs are not directly comparable: the original cRCT offered stronger control over confounding influences and larger samples, increasing power to detect small effects. Second, as in many pre-driver studies, including the original DriveFit trial, students entered the programme with already risk-averse attitudes and only a brief period elapsed between baseline and follow-up. This combination can restrict the degree of change that is detectable in the short term. Third, substituting external facilitators with classroom teachers likely introduced greater variability in delivery and reduced facilitation intensity, particularly given the absence of systematic fidelity monitoring in the present study.

A further methodological consideration relates to the timing of outcome measurement. In the original DriveFit cRCT, outcomes were assessed at baseline (T1), immediately post-intervention (T2), and again at short-term follow-up (T3; 4–6 weeks). While significant improvements were observed immediately post-intervention, several effects had diminished by the T3 follow-up. In contrast, DriveFit 2.0 was delivered across three lessons spaced over a 4–6 week period, and the present evaluation included baseline measurement and a single follow-up point after programme completion. As a result, there was no assessment point that captured immediate post-session or short-term effects following individual lessons. It is therefore possible that transient shifts in attitudes, intentions, or self-efficacy occurred during or shortly after delivery but were not detected at the later follow-up point used in this study.

It is therefore not possible to determine whether the observed differences are primarily due to delivery method, programme content, sample/contextual characteristics, or methodological factors. Nonetheless, the contrast highlights a key tension for policy and practice: teacher-led delivery enhances scalability and sustainability within PSHE and enrichment timetables, but may yield smaller or less detectable short-term effects than professionally facilitated models. Future research should aim to compare delivery models

within a single evaluative framework, ideally with robust fidelity checks, longer follow-up, and sufficient power, to isolate the contribution of facilitation versus content, while building on the demonstrated feasibility and acceptability of the teacher-led DriveFit 2.0 approach.

4.3 Strengths and limitations

The evaluation's main strengths lie in its pragmatic, school-embedded design, demonstrating that DriveFit 2.0 can be delivered within routine PSHE provision and is acceptable to both teachers and students. Outcomes were pre-specified and assessed using generally reliable multi-item scales, and the inclusion of a wait-list comparison group provided an essential counterfactual for interpreting observed changes. Several outcomes in the intervention group showed directionally positive shifts at follow-up, more so than in the comparison group, but these differences were modest and were not statistically distinct from changes in the control group. Without a comparison condition, such changes might have been misinterpreted as intervention effects, underscoring the importance of controlled designs in applied educational settings.

At the same time, several limitations constrain causal inference. The quasi-experimental (non-randomised) design limits the ability to rule out selection effects. Retention was uneven, varying by gender and some baseline driving-related characteristics, and the matched longitudinal sample was modest (n = 145), raising the possibility of response bias. The follow-up period was short and the interval between T1 and T2 varied, meaning the study may have missed two important windows of change: any immediate, short-term effects that could have been detectable directly after the lessons, and longer-term shifts that may only crystallise closer to licensing or through early real-world driving experience. This timing limitation reduces sensitivity to the types of changes DriveFit 2.0 might reasonably be expected to influence. All measures relied on self-report, with no behavioural or objective indicators, making the results vulnerable to social desirability and to biases that arise when all data come from the same type of questionnaire.

External validity is also limited: the evaluation took place in four schools/colleges within a single county, and clustering by school or class was not explicitly modelled. Implementation fidelity was not systematically assessed across all sites; aside from one observed session meaning insights relied largely on practitioner reflections. This means that conclusions about consistency of delivery must be treated cautiously. Finally, several constructs, particularly attitudes and intentions, showed high baseline scores, reflecting risk-averse starting points. These ceiling effects likely reduced the scope for detecting short-term change, a common challenge in pre-driver research.

4.4 Implications for policy and practice

The evaluation findings indicate that DriveFit 2.0 can be delivered feasibly and acceptably as a teacher-led, curriculum-embedded programme. Students generally rated the lessons as credible, useful, and engaging, and emotional responses were moderate rather than

fear-inducing, consistent with PSHE's strengths-based, non-threatening ethos. These are encouraging indicators for scalability within school and college settings.

However, evidence of impact on psychological outcomes was limited. Aside from a small increase in perceived risk, no significant intervention effects were found for attitudes, intentions, or self-efficacy once baseline levels were taken into account. There were hints of improvement on several measures among intervention students, but these improvements were modest and not statistically distinct from changes seen in the comparison group once baseline levels were taken into account. This means that, in this study, DriveFit 2.0 did not demonstrate measurable short-term effects on key determinants of young driver safety.

The weaker effects compared with the original cRCT of DriveFit likely reflect a combination of factors:

- the less intensive, teacher-led delivery model,
- moderate and variable fidelity,
- high baseline scores (reducing scope for detectable change)
- delivery by familiar classroom teachers rather than external facilitators, potentially reducing perceived novelty or salience, and
- the quasi-experimental design and short, heterogeneous follow-up window.

A cluster randomised trial would have offered stronger causal evidence but was beyond the scope of this project, especially given the practical challenges of recruiting large numbers of schools within one term. The quasi-experimental approach used here was a pragmatic choice, particularly given that the original DriveFit programme had already been tested in a cRCT. Accordingly, this study was intended to assess feasibility, acceptability, and short-term patterns of change rather than to provide definitive evidence of impact, so conclusions about effectiveness should remain cautious.

From a policy perspective, these findings suggest that teacher-delivered models may hold promise as part of a wider youth road safety strategy, but that delivery mode matters. The contrast between the outcomes of DriveFit 2.0 and the original programme raises questions about whether a purely teacher-led approach is sufficient to achieve behaviourally meaningful change. Hybrid models, combining teacher delivery with periodic professional facilitation, guided debriefs, or digital reinforcement, may strike a stronger balance between scalability and impact.

If DriveFit 2.0 is to be deployed more widely, implementation should be supported through:

- concise teacher training and induction,
- reliable technology and resource access,
- simple fidelity tools or checklists to promote consistent delivery, and
- opportunities for reflective follow-up or supplementary reinforcement.

Finally, further evaluation is warranted. More robust designs, ideally a larger-scale controlled trial with fidelity monitoring, and longer-term follow-up into early driving, would help determine whether DriveFit 2.0 can produce meaningful, sustained improvements in young driver safety.

4.5 Recommendations for programme refinement

While feasibility and acceptability were demonstrated, several refinements could enhance the potential impact of DriveFit 2.0. Elements that explicitly support realistic risk awareness and judgement, such as scenario-based discussions, peer-influence activities, and commitment-setting tasks, should be retained and potentially strengthened, as these align with the modest increase in perceived risk observed in the evaluation.

Because negative emotional responses were moderate and did not indicate distress, there is also scope to enrich experiential components (for example, through more detailed scenarios or immersive decision-making tasks) while remaining firmly within a non–fear-based, PSHE-appropriate approach. Teacher delivery could be supported further through targeted facilitation guidance, practical troubleshooting strategies for common barriers (such as IT issues), and clear fidelity tools, including session checklists or delivery flowcharts. Strengthening mechanisms for monitoring fidelity across settings would also help ensure that variations in delivery are understood and accounted for in future evaluations.

Importantly, any refinements should be implemented alongside formal evaluation rather than assumed to improve outcomes, to ensure that changes enhance both feasibility and impact.

4.6 Recommendations for future evaluation

Future evaluations should build on insights from both the original DriveFit cRCT and the present study. A key priority is to compare delivery models directly, teacher-led, externally facilitated, or hybrid approaches, to determine whether differences in facilitation intensity account for variation in effectiveness. While the quasi-experimental design used here was appropriate for a resource-constrained study, more robust methodologies are required to establish impact conclusively. Well-powered randomised controlled or cluster-randomised trials would enable modelling of school- and class-level clustering, reduce potential bias, and strengthen causal inferences.

Longer-term follow-up is also essential. Impact may emerge at different points: some effects may appear immediately post-intervention but fade within weeks, while others, such as shifts in risk perception or self-efficacy, may develop gradually as young people begin supervised practice or transition into early licensure. Evaluations should therefore consider multiple follow-up points spanning this period.

Incorporating objective or proxy behavioural indicators (e.g., telematics, simulated driving tasks, parent reports) would also help address the limitations of self-report measures and provide richer evidence of behavioural relevance. Finally, future studies should include robust retention strategies and reduce the risk of biased follow-up samples.

Strengthening methodological design, broadening outcome measures, and explicitly testing delivery model effects will help future research build on these findings and deepen understanding of DriveFit 2.0's potential role in improving young driver safety.

5. Conclusion

The evaluation of DriveFit 2.0 shows that a teacher-led, curriculum-embedded model of pre-driver education is both feasible and acceptable within schools and colleges. The programme was delivered within routine PSHE/enrichment provision, generally followed its intended three-lesson structure, and was viewed by students as credible, useful, and moderately engaging. Emotional responses were neutral-to-moderate rather than fear-inducing, indicating that a strengths-based, non-shock approach to pre-driver education can be integrated into PSHE at scale in a way that aligns with best practice.

Short-term psychological impacts were more limited. After adjusting for baseline levels and comparison-group trends, DriveFit 2.0 did not produce significant changes in attitudes, intentions, or self-efficacy across mobile phone use, fatigue, or speeding, in a sample that was already relatively risk-averse at baseline. The only clear quantitative effect was a small but statistically significant increase in perceived risk, with intervention students rating risky driving scenarios as slightly more dangerous at follow-up than controls. Several other outcomes moved in a direction consistent with safer behaviour among intervention students, but these changes were modest and not

statistically distinct from those seen in the comparison group once initial scores were taken into account. Given the quasi-experimental design, modest longitudinal sample, short and variable follow-up period, reliance on self-report, and lack of systematic fidelity monitoring, these findings should be interpreted cautiously.

Set alongside the original DriveFit cRCT, which found broader positive effects under a professionally facilitated model, DriveFit 2.0 appears to yield weaker short-term psychological change under teacher-led delivery. The present evaluation cannot disentangle whether this reflects differences in facilitation intensity, programme content, sample/context, or methodological factors such as power and the timing of outcome measurement. In particular, the absence of immediate post-session measurement means that any short-lived effects occurring during or shortly after delivery may not have been captured. Nonetheless, the contrast highlights a central tension for policy and practice: teacher-led models offer scalability and sustainability, but may require additional supports (for example, enhanced experiential content, structured risk-appraisal tasks, and clearer facilitation guidance) or hybrid delivery models if stronger behavioural determinants are to be influenced.

The process findings are encouraging in terms of reach and engagement. DriveFit 2.0 was delivered with moderate-to-good fidelity in varied post-16 settings, and acceptability ratings were broadly similar across gender categories. Although the intervention group was predominantly male, descriptive trends across several outcome measures, including perceived risk, tended to move in a more safety-supportive direction among intervention students than controls. These patterns were modest and not statistically robust (apart from for perceived risk), but they indicate that DriveFit 2.0 was associated with favourable directional shifts within a sample traditionally considered at elevated crash risk.

Overall, this evaluation provides a strong foundation for understanding the feasibility, reception, and early effects of DriveFit 2.0 in schools and colleges. It demonstrates that DriveFit 2.0 can be embedded in school/college practice, identifies a small but consistent effect on perceived risk, and highlights key constraints and design challenges that future research must address. Future work should focus on strengthening the programme's support for realistic risk awareness and judgement and its experiential learning elements, alongside clearer support and monitoring tools for teachers, and further evaluation over longer timeframes and across different delivery models. This will help build a stronger evidence base on how DriveFit 2.0 and comparable interventions can most effectively support safer driving behaviours among young people.

6. References

Ajzen, I. (1991). The Theory of Planned Behaviour. *Organizational Behavior and Human Decision Processes, 50*, 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Box, E., & Dorn, L. (2023). A cluster randomised controlled trial (cRCT) evaluation of a pre-driver education intervention using the Theory of Planned Behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*, *94*, 379–397. https://doi.org/10.1016/j.trf.2023.03.001

Brewster, S. E., Elliott, M. A., & Kelly, S. W. (2015). Evidence that implementation intentions reduce drivers' speeding behavior: Testing a new intervention to change driver behavior. *Accident Analysis and Prevention*, *74*, 229–242. https://doi.org/10.1016/j.aap.2014.11.006

Conner, M., & Sparks, P. (2005). Theory of Planned Behaviour and Health Behaviour. In M. Conner & P. Norman (Eds.), *Predicting Health Behaviour* (2nd Edition, pp. 170–222). Open University Press.

Cuenen, A., Brijs, K., Brijs, T., Van Vlierden, K., Daniels, S., & Wets, G. (2016). Effect evaluation of a road safety education program based on victim testimonials in high schools in Belgium. *Accident Analysis & Prevention*, *94*, 18–27. https://doi.org/10.1016/j.aap.2016.05.006

Glendon, A. I., McNally, B., Jarvis, A., Chalmers, S. L., & Salisbury, R. L. (2014). Evaluating a novice driver and pre-driver road safety intervention. *Accident Analysis & Prevention*, *64*, 100–110. https://doi.org/10.1016/j.aap.2013.11.017

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*(7), 493–503. https://doi.org/10.1037/0003-066X.54.7.493

Ivers, R., Senserrick, T., Boufous, S., Stevenson, M., Chen, H. Y., Woodward, M., & Norton, R. (2009). Novice drivers' risky driving behavior, risk perception, and crash risk: Findings from the DRIVE study. *American Journal of Public Health*, *99*(9), 1638–1644. https://doi.org/10.2105/AJPH.2008.150367

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, *46*(1), 81–95. https://doi.org/10.1007/s12160-013-9486-6

Road Safety Analysis. (2015). *Safe Drive Stay Alive Evaluation Greater Manchester and Surrey*.

Rowe, R., Andrews, E., Harris, P. R., Armitage, C. J., McKenna, F. P., & Norman, P. (2016). Identifying beliefs underlying pre-drivers' intentions to take risks: An application of the Theory of Planned Behaviour. *Accident Analysis & Prevention*, *89*, 49–56. https://doi.org/10.1016/j.aap.2015.12.024

World Health Organisation. (2023). Global Status Report on Road Safety 2023. WHO. https://www.who.int/publications/i/item/9789240086517

Annex A: Survey questions

PRE SURVEY		
No.	**Question**	**Response options**
About you		
Q1	What gender do you identify as?+	Male, Female, Non-binary, Prefer not to say
Q2	How old are you?+	15, 16, 17, 18, 19
Q3	What is your ethnic group+	White; Mixed or multiple ethnic groups; Asian or Asian British; Black, African, Caribbean or Black British; Other ethnic group; Prefer not to say
Q4	What school/college do you attend?	[Selection of all schools/colleges taking part in the trial]
Q5	What is your name?	First name, surname
About learning to drive		
Q6	Do you plan to learn to drive?	Yes, I have already passed my driving test (Please state MM/YYYY of test pass); Yes, I am currently learning; Yes, in the next 12 months; Yes, in the next 5 years; Maybe, at some point; No, never.
Q7	In total, how many cars or vans are owned, or available for use, by members of your household?	None, 1, 2, 3, 4, 5 or more (Please write in the number)
Q8	How many hours of driving practice do you think are necessary before taking the driving test?	0-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, 100+
Q9	How often do you/do you think you will practice driving each week whilst learning to drive?	Less than once a week, 1-2 times a week, 3-4 times a week, 5-6 times a week, everyday
Q10	How often do you (or do you think you will) practice driving in different environments (e.g. city, highways, rural areas etc.)	Never, very rarely, rarely, sometimes, often, very often, always
Q11	How confident are you in your ability to identify potential hazards while driving?	Not at all confident–completely confident [7-point scale]
Q12	To what extent do you agree or disagree with the following statement: When driving or riding as a passenger, I actively look for potential hazards	Strongly agree – Strongly disagree + [7-point scale]

PRE SURVEY		
No.	**Question**	**Response options**
Your attitudes and views about specific driving behaviours		
Q13	Driving whilst messaging or talking on a hand-held mobile is… [1]	Matrix response Harmful – Beneficial+, Negative – Positive+, Wise – Foolish, Pleasant – Unpleasant [7-point scale]
Q14	How likely and willing would you be to drive whilst messaging or talking on a hand-held mobile phone? [1]	Matrix response Likely – Unlikely, Very Willing – Not very willing [7-point scale]
No.	**Question**	**Response options**
Q15	Driving whilst feeling very tired is… [1]	Matrix response Harmful – Beneficial+, Negative – Positive+, Wise – Foolish, Pleasant – Unpleasant [7-point scale]
Your attitudes and views about general driving behaviours		
Q16	How likely and willing would you be to drive whilst feeling very tired? [1]	Matrix response Likely – Unlikely, Very Willing – Not very willing [7-point scale]
Q17	Driving over the speed limit is… [1]	Matrix response Harmful – Beneficial+, Negative – Positive+, Wise – Foolish, Pleasant – Unpleasant [7-point scale]
Q18	How likely and willing would you be to drive over the speed limit? [1]	Matrix response Likely – Unlikely, Very Willing – Not very willing [7-point scale]
Q19	When driving, how safe do you think the following situations are? [2]	Matrix response: Driving with 2 or more passengers, Driving between midnight and 6am, Driving at 70mph in a 60mph zone, Driving at 40mph in a 30mph zone, Driving while talking on a mobile phone, Driving a car which is over 10 years old, Driving with a blood alcohol level just above the legal limit, Driving whilst messaging on a mobile phone, Driving after smoking marijuana, Going through a red light Response options: Always safe, Mostly safe, Sometimes safe, Rarely safe [4-point scale]
Q20	To what extent do you agree or disagree with the following statement: As a driver, I am confident that I will meet the challenge of maintaining safe driving behaviours and managing my personal risk whilst driving.	Strongly agree – Strongly disagree + [5-point scale]

POST SURVEY		
No.	**Question**	**Response options**
About you		
Q1	What school/college do you attend?	[Selection of all schools/colleges taking part in the trial]
Q2	What is your name?	First name, surname
About learning to drive		
Q3	Do you plan to learn to drive?	Yes, I have already passed my driving test (Please state MM/YYYY of test pass); Yes, I am currently learning; Yes, in the next 12 months; Yes, in the next 5 years; Maybe, at some point; No, never.
Q4	How many hours of driving practice do you think are necessary before taking the driving test?	0-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, 100+
Q5	How often do you/do you think you will practice driving each week whilst learning to drive?	Less than once a week, 1-2 times a week, 3-4 times a week, 5-6 times a week, everyday
Q6	How often do you (or do you think you will) practice driving in different environments (e.g. city, highways, rural areas etc.)	Never, very rarely, rarely, sometimes, often, very often, always
Q7	How confident are you in your ability to identify potential hazards while driving?	Not at all confident–completely confident [7-point scale]
Q8	To what extent do you agree or disagree with the following statement: When driving or riding as a passenger, I actively look for potential hazards	Strongly agree – Strongly disagree + [7-point scale]
Your attitudes and views about specific driving behaviours		
Q9	Driving whilst messaging or talking on a hand-held mobile is… [1]	Matrix response Harmful – Beneficial +, Negative – Positive +, Wise – Foolish, Pleasant – Unpleasant [7-point scale]
Q10	How likely and willing would you be to drive whilst messaging or talking on a hand-held mobile phone? [1]	Matrix response Likely – Unlikely, Very Willing – Not very willing [7-point scale]
Q11	Driving whilst feeling very tired is… [1]	Matrix response Harmful – Beneficial +, Negative – Positive +, Wise – Foolish, Pleasant – Unpleasant [7-point scale]
Q12	How likely and willing would you be to drive whilst feeling very tired? [1]	Matrix response Likely – Unlikely, Very Willing – Not very willing [7-point scale]
Q13	Driving over the speed limit is… [1]	Matrix response Harmful – Beneficial +, Negative – Positive +, Wise – Foolish, Pleasant – Unpleasant [7-point scale]

POST SURVEY		
No.	**Question**	**Response options**
No.	**Question**	**Response options**
Q14	How likely and willing would you be to drive over the speed limit? [1]	Matrix response Likely – Unlikely, Very Willing – Not very willing [7-point scale]
Q15	When driving, how safe do you think the following situations are? [2]	Matrix response: Driving with 2 or more passengers, Driving between midnight and 6am, Driving at 70mph in a 60mph zone, Driving at 40mph in a 30mph zone, Driving while talking on a mobile phone, Driving a car which is over 10 years old, Driving with a blood alcohol level just above the legal limit, Driving whilst messaging on a mobile phone, Driving after smoking marijuana, Going through a red light Response options: Always safe, Mostly safe, Sometimes safe, Rarely safe [4-point scale]
Q16	To what extent do you agree or disagree with the following statement: As a driver, I am confident that I will meet the challenge of maintaining safe driving behaviours and managing my personal risk whilst driving.	Strongly agree – Strongly disagree + [5-point scale]
Process evaluation questions (Intervention group only)		
Q17	To what extent do you agree or disagree that the DriveFit 2.0 lessons were… [3]	Matrix response: Credible, Useful, Interesting, Important, Informative, Enjoyable, Boring, Surprising, Shocking, Worrying, Frightening Response options: Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree + [5-point scale]
Q18	To what extent do you agree or disagree that the DriveFit 2.0 lessons… [4]	Matrix response: Have provided you with new insights, Have been beneficial to you personally, Have covered ideas that you see yourself doing, Have highlighted changes that you need to make, Will make a difference to your decisions and behaviours as a driver or passenger Response options: Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree+ [5-point scale]
Q19	If you have any further comments to make about DriveFit 2.0 please write them here.	[Open text response]

Nb. Higher scores indicate safer behaviours. Reverse coded survey items are marked with a +.
1 Measured using adapted standard measures for Theory of Planned Behaviour components (Conner & Sparks, 2005; Rowe et al., 2016), 2 an adapted perceptions of risk scale (Glendon et al., 2014; Ivers et al., 2009) 3 Cognitive response measured by measures from Cuenen et al. (2016) 4 Face validity measured by adapting measures from Road Safety Analysis (2015)

Annex B: Process evaluation questions for DriveFit 2.0 deliverers

Fidelity: Implementation consistency

Objective: Understand how closely delivery matched the intended design

1. To what extent do you think the DriveFit 2.0 sessions as planned (e.g., in terms of structure, content, sequence)?
2. Were there any components of the intervention (lessons, activities, films) that were routinely skipped or adapted? If so, why?
3. Do you think teachers felt adequately supported to deliver the intervention as designed? Did the teachers differ in their background and skills?
4. How did students respond to the planned materials and structure – do you think this effected how it was delivered
5. Overall, from your discussions with teachers, do you think the sessions were delivered as intended – score 1-10 (1 = not at all, 10 = completely)

Dose delivered: What was delivered

Objective: Determine the completeness of intervention delivery

6. How many DriveFit 2.0 sessions were delivered in each school/college?
7. Were there any challenges in ensuring that all three sessions were completed in full?
8. How did you track what was delivered, and was this documentation (e.g., logs, reflections) manageable and accurate?

Dose received: Engagement and satisfaction

Objective: Assess engagement and perceived value by all stakeholders

9. How engaged were the students during the sessions (e.g., participation, attention, feedback)?
10. What aspects of the intervention do you think students found most or least enjoyable or relevant?
11. Were you personally satisfied with the quality and impact of the materials you saw delivered? Why or why not?
12. What feedback (formal or informal) did you receive from teachers or school staff about the intervention?

Reach: Who received the intervention

Objective: Understand how widely the intervention was accessed

13. Did you have any insight into how many students from each school/college actually participated in the sessions?
14. Were there any patterns in attendance or engagement (e.g., certain student groups more or less likely to attend or engage)?

Recruitment: How participants were enrolled

Objective: Understand how schools and practitioners were recruited

15. Can you describe how you approached/selected schools/colleges to deliver DriveFit 2.0?
16. What worked well in the recruitment process (of schools, students)?
17. Were there any barriers that delayed or complicated getting started at each site?

Context: Barriers and facilitators to delivery

Objective: Identify environmental and contextual factors influencing delivery

18. What factors helped deliver the intervention effectively (e.g., school culture, facilities, timing)?
19. What barriers did you encounter (e.g., scheduling issues, student behaviour, staff support)?
20. Did any unexpected events or contextual factors influence how the programme was delivered?
21. What recommendations would you give for improving delivery in future settings or cohorts?

Annex C: Descriptive statistics for practice and hazard perception

Table E-1. Descriptive statistics (means and standard errors) for practice and hazard perception outcomes at baseline (T1) and follow-up (T2), Intervention vs. Control

Outcome	Group	T1 Mean (SE)	T2 Mean (SE)	Δ (T2–T1)
Expected hours to pass test	Control (n=65)	4.37 (0.30)	4.25 (0.31)	–0.12
	Intervention (n=80)	4.00 (0.30)	4.13 (0.30)	+0.13
Expected weekly frequency	Control	2.42 (0.08)	2.29 (0.08)	–0.13
	Intervention	2.28 (0.09)	2.36 (0.09)	+0.08
Expected variety of environments	Control	4.26 (0.16)	4.38 (0.14)	+0.12
	Intervention	3.96 (0.12)	4.09 (0.14)	+0.13
Confidence in identifying hazards	Control	4.85 (0.19)	2.78 (0.16)	–2.07
	Intervention	4.79 (0.15)	2.71 (0.14)	–2.08
Active hazard scanning (“I actively look for potential hazards”)	Control	3.66 (0.25)	4.20 (0.25)	+0.54
	Intervention	4.35 (0.21)	3.84 (0.20)	–0.51

Note. Δ (T2–T1) represents the change in mean scores between baseline and follow-up. Positive values (shown in green) indicate increases (e.g., greater practice expectations or stronger hazard perception), while negative values (shown in red) indicate decreases. These descriptive changes should be interpreted with caution, as ANCOVA analyses (see main text) found no statistically significant group effects.

Annex D: Descriptive statistics for attitudes and intentions across risk domains (mobile phone use, fatigue, speeding)

Table F-1: Descriptive statistics (means and standard errors) for attitudes and intentions across risk domains (mobile phone use, fatigue, speeding) at baseline (T1) and follow-up (T2), Intervention vs. Control

Domain	Outcome	Group	T1 Mean (SE)	T2 Mean (SE)	Δ (T2–T1)
Mobile phone	Attitudes	Control (n=65)	6.50 (0.12)	6.17 (0.14)	–0.33
		Intervention (n=80)	5.96 (0.15)	6.04 (0.13)	+0.08
	Intentions	Control	6.15 (0.16)	6.11 (0.17)	–0.04
		Intervention	5.74 (0.18)	5.79 (0.16)	+0.05
Fatigue	Attitudes	Control	6.31 (0.13)	6.17 (0.12)	–0.14
		Intervention	6.18 (0.13)	6.19 (0.13)	+0.02
	Intentions	Control	5.73 (0.18)	5.92 (0.16)	+0.19
		Intervention	5.62 (0.17)	5.88 (0.15)	+0.26
Speeding	Attitudes	Control	6.08 (0.16)	5.97 (0.16)	–0.11
		Intervention	5.78 (0.16)	5.81 (0.16)	+0.03
	Intentions	Control	5.74 (0.19)	5.78 (0.19)	+0.04
		Intervention	5.43 (0.19)	5.64 (0.17)	+0.21

Note. Δ (T2–T1) represents the change in mean scores between baseline (T1) and follow-up (T2). Positive values (shown in green) indicate an increase in risk-averse attitudes or intentions over time, while negative values (shown in red) indicate a reduction. These descriptive shifts should be interpreted with caution, as ANCOVA analyses showed that none of the group differences were statistically significant.

Annex E: Descriptive statistics for self-efficacy and perceived risk

Table G-1: Descriptive statistics (means and standard errors) for perceived risk and self-efficacy at baseline (T1) and follow-up (T2), Intervention vs. Control

Domain	Group	T1 Mean (SE)	T2 Mean (SE)	Δ (T2–T1)
Self-efficacy	Control (n=65)	3.89 (0.17)	3.89 (0.17)	0.00
	Intervention (n=80)	3.84 (0.13)	3.89 (0.13)	+0.05
Perceived risk	Control	3.23 (0.05)	3.23 (0.06)	0.00
	Intervention	3.16 (0.06)	3.28 (0.05)	+0.12

Note. Δ (T2–T1) represents the change in mean scores between baseline and follow-up. Positive values (shown in green) indicate increases (e.g., greater confidence or heightened risk perception), no change values (shown in orange) indicate no change and negative values (red) indicate decreases. These descriptive changes should be interpreted with caution, as ANCOVA analyses (see main text) did not show statistically significant group effects for self-efficacy. Statistically significant group effects were found for perceived risk.

RAC Foundation

Mobility · Safety · Economy · Environment

The Royal Automobile Club Foundation for Motoring Ltd is a transport policy and research organisation which explores the economic, mobility, safety and environmental issues relating to roads and their users. The Foundation publishes independent and authoritative research with which it promotes informed debate and advocates policy in the interest of the responsible motorist.

RAC Foundation
89–91 Pall Mall
London
SW1Y 5HS

Tel no: 020 7747 3445
www.racfoundation.org

Registered Charity No. 1002705

Designed and printed by
The Javelin Partnership Ltd
Tel: 0118 907 3494