**nfer**
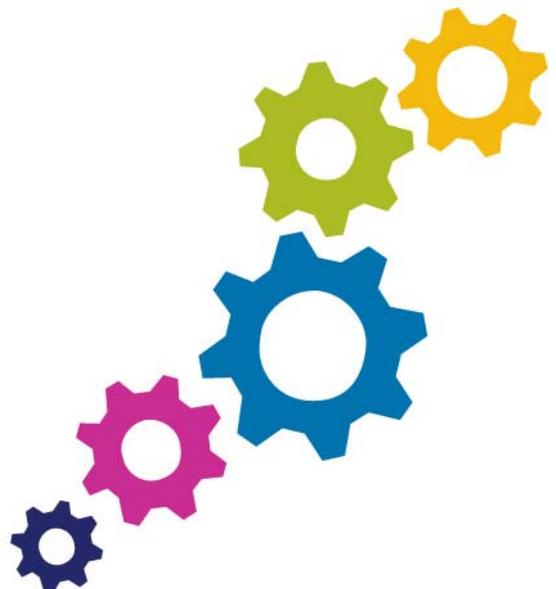
**Evidence for Excellence in Education**

# Research into the Impact of Bikeability Training on Children's Hazard Perception and Behaviour

National Foundation for Educational Research (NFER)

● independent ● insights ● breadth ● connections ● outcomes

# Research into the impact of Bikeability training on children's hazard perception and behaviour

Claire Hodgson
Jack Worth

# Contents

# Project team

| | |
|---|---|
| Catherine Kirkup | Researchers |
| Claire Hodgson | |
| Michael Frearson | |
| Nicola Keogh | Project Co-ordinator |
| Jack Worth | Statistician |
| Ian Brownlie | Research and Product Operations |

# Acknowledgements

# Executive summary

## Context

Bikeability (2014) is described as 'cycling proficiency for the 21st century'. The training is practical, skill-based, outcome-led and designed to 'boost the confidence of the trainee and to minimise risk'. There are three levels of training and children typically start Bikeability lessons once they have learnt to ride a bike. Level 2 training is typically provided to children in Year 5 or 6 before they leave primary school. The policy purpose of Bikeability is to give children the skills and confidence needed to cycle on today's roads and so encourage more people to cycle more often with less risk.

It is against this background that this research was designed to test the hypothesis that Bikeability training improves a child's ability to assess and respond appropriately to on-road hazards faced by people who cycle. It sought evidence to confirm whether or not Bikeability-trained children are less likely to be involved in collisions or other road accidents, and thus less at risk of injury than untrained children, when cycling on the road.

The research, undertaken by the National Foundation for Educational Research (NFER), was commissioned by The Bikeability Support Team at Steer Davies Gleave with funding from the Department for Transport.

NFER is the UK's largest independent provider of research, assessment and information for education, training and children's services. NFER is well known for its well established relationship with schools in the UK, as well as in-depth experience, knowledge and expertise in the field of assessment having produced high-stakes national tests and assessments over the last 50 years.

## The research question

The main research question for the study was: how does Bikeability affect the ability of children to perceive and appropriately respond to hazards when cycling on the road, if at all?

This question was explored by means of an on-screen quiz to test knowledge and skills of hazard perception and risk mitigation. The quiz was taken by both trained and untrained pupils and validated by a practical on-road assessment of Bikeability-trained children.

## Key findings

- Children who participated in Bikeability Level 2 training scored significantly higher on the hazard perception and risk mitigation quiz than children who had not received training.

- The effect of the Bikeability Level 2 training was undiminished when children re-took the quiz more than two months after training. This suggests that the association between training and increased hazard perception and risk mitigation strategies was sustained.

- The difference in scores, referred to as 'hazard perception and risk mitigation ability', for children who had participated in training (trained) and those who had not (comparison) is shown below.

**Comparison of mean pupil ability scores at baseline before training (phase 1) and immediately after\* training (phase 2)**
**\*within 1-3 weeks of training**

Hazard perception and risk mitigation ability
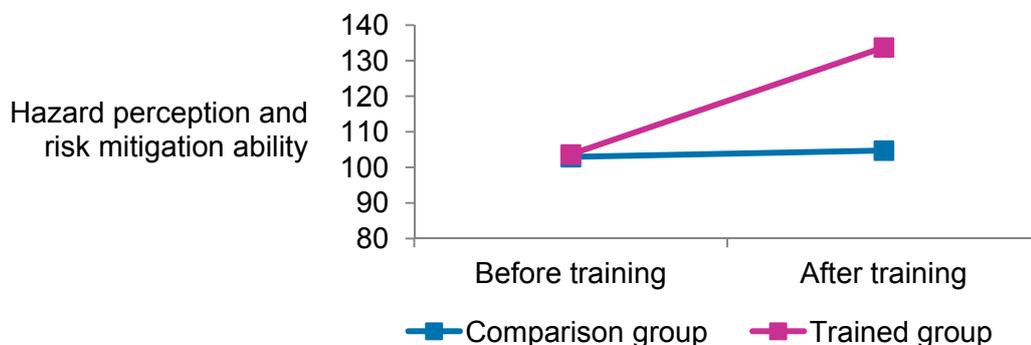


**Comparison of mean pupil ability scores at baseline before training (phase 1) and at least two months\* after training (phase 3)**
**\*within 2-3 months of training**
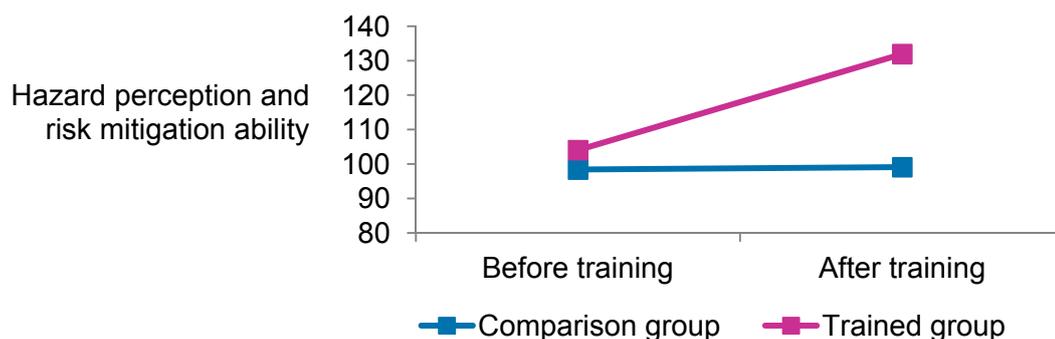
Hazard perception and risk mitigation ability



- The size of the association between training and hazard perception is very large, with an effect size of 1.6. The change in performance for children who had participated in training (trained) and those who had not (comparison) is shown in the score distribution chart below.



'Hazard perception and risk mitigation ability' score
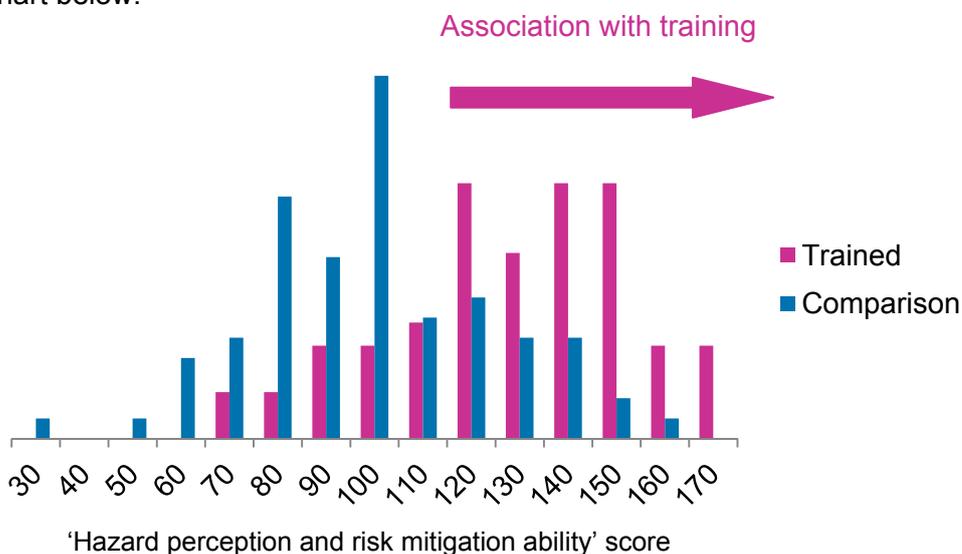
- Children who participated in training reported increased confidence when cycling on the road compared to their initial level of confidence. This increase was statistically significant. There was no increase in confidence in the comparison group.

- There was no association between training and frequency of cycling – children did not report that they cycle more often as a result of receiving Bikeability training, despite the fact that they had increased confidence.

- In the on-screen quiz, across all three phases, 'observation' was the highest scoring domain. The largest gains associated with training in the on-screen quiz were in the domains of 'road position' and 'priorities'.

- In the practical assessments, in both phases (2 and 3) 'observation' was the highest scoring domain whilst 'communication' was the lowest scoring.

- There was a significant decrease in the mean scores achieved on the practical assessment between phase 2 and phase 3. This suggest that whilst trained children achieved higher scores for the on-screen quiz and sustained this over a period of time, the ability to put that knowledge into practice can decline over time if the skills are not practised.

- The correlation between the practical assessment and the on-screen quiz was positive and statistically significant. This confirms that the practical assessment does validate the on-screen quiz. Further, it suggests that the on-screen quiz provides a reliable measure of aspects of the Bikeability Level 2 training and assessment.

- The on-screen quiz functioned appropriately with a reliability measure of 0.76 (Cronbach's Alpha) indicating that it discriminates well between pupils who achieve higher and lower hazard perception and risk mitigation scores.

## Methodology

### Who was in the sample?

The research involved pupils who were in Year 5 in summer 2014 and tracked them as they moved into Year 6 in the autumn term.

In total, 29 schools were involved in the study, with six of those schools participating in the practical assessments. A total of 668 pupils were involved in taking one or more on-screen quizzes.

Participating schools and their pupils were either in the intervention or comparison group. Schools in the intervention group had pupils who participated in Bikeability training during the summer term (trained pupils). Pupils in the comparison schools did not receive any training in the summer term.

### When did the data collection take place?

There were three data collection points:

- Phase 1 – baseline prior to any training taking place (summer term 2014)

- Phase 2 – within one to three weeks after the completion of training (summer term 2014)

- Phase 3 – two to three months after the completion of training (autumn term 2014)

An on-screen quiz was completed by pupils at each phase.

The practical assessments were taken at phase 2 and phase 3 by pupils who had successfully completed Bikeability Level 2 training.

## What did the assessment involve?

### *On-screen quiz*

NFER developed an on-screen quiz designed to assess four domains that underpin effective hazard perception and risk mitigation strategies: observation, communication, road position and priorities. In order to engage respondents, the quiz told the story of three children's cycling journeys. This allowed for inclusion of photographs and film clips showing different aspects of the children's journeys, for example, choosing where and when to start their ride, considering road position and priorities for different manoeuvres and completing the journey.

The purpose of the on-screen quiz was two-fold: to measure pupils' hazard perception and risk mitigation ability and to establish the functioning of a variety of questions about hazard perception and risk mitigation with a view to providing a pool of questions for potential future use.

### *Practical assessment*

In order to ascertain whether or not the on-screen assessment was a reliable tool for measuring hazard perception and risk mitigation ability, some children who had passed their Bikeability Level 2 training were also given a practical assessment. The practical assessment was carried out by qualified and experienced National Standard Instructors (NSIQs) and involved pupils in completing two drills. These were designed to provide sufficient opportunities to demonstrate competence, confidence and consistency in the four domains also covered by the on-screen quiz. The scores achieved by pupils on the on-screen quiz and practical assessment were analysed to establish if there was a correlation.

## Recommendations

The on-screen hazard perception quiz could be used for a number of purposes to support the delivery and development of Bikeability training including:

- monitoring the effectiveness of the training and for identifying any particular areas which may need developing or strengthening

- providing a way of evaluating and moderating the training given by different providers across the country

- providing a way of evaluating and moderating the different models of training

- monitoring the impact of the training over a longer period of time to help identify which domains are sustained and if there are any areas for which follow-up or refresher training may be usefully implemented.

# 1 Introduction

## 1.1 Background

Bikeability aims to encourage everyday cycling by developing the skills, knowledge and understanding needed for effective and confident on-road cycling. More than a million children have participated in Bikeability since its launch in 2007; currently, about half of all children are trained before they leave primary school, supported by annual funding of £11m from the Department for Transport (DfT).

The policy purpose of Bikeability is to 'get more people cycling more safely, more often'. The Bikeability Support Team at Steer Davies Gleave commissioned NFER to investigate whether Bikeability training does improve a child's ability to assess and appropriately respond to on-road cycling hazards. The resulting evidence may suggest that Bikeability-trained children will be less likely to be involved in collisions or other road accidents, and thus cycle 'more safely' than untrained children.

Most Bikeability training occurs during school time in Years 5 and 6 and combines Bikeability Level 1 (developing excellent bicycle handling skills in traffic-free environments and preparing for on-road cycling) and Level 2 (cycling on single-lane roads and using junctions). Achievement of the National Standard for Cycle Training outcomes, which underpin Bikeability at Level 2, certify a trainee's ability to demonstrate consistently, competently and confidently independent decision making, sound hazard perception and safe cycling strategies.

## 1.2 Aims and objectives

The main objective of the research was to test the hypothesis that Bikeability training improves a child's ability to assess and respond appropriately to on-road hazards faced by people who cycle.

The research tested this overall hypothesis but also aimed to discover what particular aspects of children's ability to perceive and respond appropriately to such hazards are improved by Bikeability training, relative to untrained children.

In addition to answering the key research question, our research therefore enables conclusions to be drawn about the strengths of Bikeability training and any recommendations as to how such training might be improved.

## 1.3 Research questions

The main research question for the study was: how does Bikeability affect the ability of children to perceive and respond appropriately to hazards when cycling on the road, if at all?

Within this, a set of supplementary research questions were explored to address how (if at all) such effects are achieved:

- Are trained children more aware of common cycling hazards than untrained children - i.e. are they better at hazard perception and risk mitigation than untrained children?

- Are trained children better at making independent decisions which reduce risk?

- Do trained children select safer cycling strategies – i.e. observation, communication, road position, priorities?

  o Observation - are they more aware of other road users?

  o Communication - are they better at knowing when to signal their intentions (and when not), and how to communicate to best effect?

  o Road position - do they select optimal road positions for different phases of their cycling journey?

  o Priorities - do they have a better understanding of their priorities and those of other road users?

The research questions were explored by means of an on-screen quiz that assesses children's hazard perception and risk mitigation ability. The quiz was taken by both trained and untrained pupils and validated by a practical assessment of Bikeability-trained children.

## 1.4 Defining hazard perception

The Bikeability Delivery Guide defines hazard perception as 'the ability to identify hazards ahead well in advance thereby enabling the cyclist to anticipate, prepare for and reduce their risk' (p. 48).

At Level 2, one of the compulsory outcomes is 'be aware of potential hazards'. Good observation improves hazard perception and thus allows for good forward planning. Preparation for hazards helps to reduce risk.

In particular, awareness of potential hazards refers to:

- demonstrating an awareness of other road users at all times, both in front and behind

- looking for hazards

- being aware of pedestrians and others on the pavement ahead of them, who might step into their path, and of driveways and other entrances, from which vehicles might emerge into their path.

Four elements of effective hazard perception and risk mitigation strategies - observation, communication, road position, priorities - permeate most outcomes at Level 2. These form

the focus of the questions in the on-screen quiz and practical assessment described in sections 2.3 and 2.4 respectively.

# 2 Methodology

## Key Points

- The study ran from summer term 2014 to autumn term 2014 and had three data collection points:

  Phase 1 – baseline prior to any training taking place (summer term 2014)

  Phase 2 – within 1-3 weeks after the completion of training (summer term 2014)

  Phase 3 – 2-3 months after the completion of training (autumn term 2014).

- It involved pupils who were in Year 5 at the start of the study and tracked them as they moved into Year 6 in the following autumn term.

- Intervention group – pupils in schools who participated in Bikeability training during the study (trained pupils).

- Comparison group – pupils in schools who did not receive Bikeability training during the study (untrained pupils).

- An on-screen quiz was completed by pupils from both groups at each phase.

- The practical assessments were taken at phase 2 and phase 3 by pupils from the intervention group who had successfully completed Bikeability Level 2 training.

## 2.1 Research design

We considered the various options for providing evidence of children's ability to perceive road hazards and deploy risk mitigation strategies and felt that in order to be a valid assessment of hazard perception and response, the assessment should comprise practical and theoretical components.

The research design, summarised overleaf, comprised:

- an analysis of relevant collision and injury data and literature relating to cycle training for children to inform the development of the on-screen quiz questions

- recruitment of Year 5 pupils via liaison with Bikeability schemes, schools and parents

- the development of an appropriate, accessible and engaging on-screen quiz to assess hazard perception and risk mitigation ability

- a realistic, safe and credible practical assessment of the hazard perception of Bikeability trained pupils, administered by qualified and experienced National Standard Instructors (NSIQs)

- data collection and analysis, including a comparison of the results of the on-screen quiz taken by trained (intervention) and untrained (comparison) children at three data collection points – phase 1 (baseline), phase 2 (1-3 weeks after training) and phase 3 (2-3 months after training)

- reporting of key findings and recommendations.
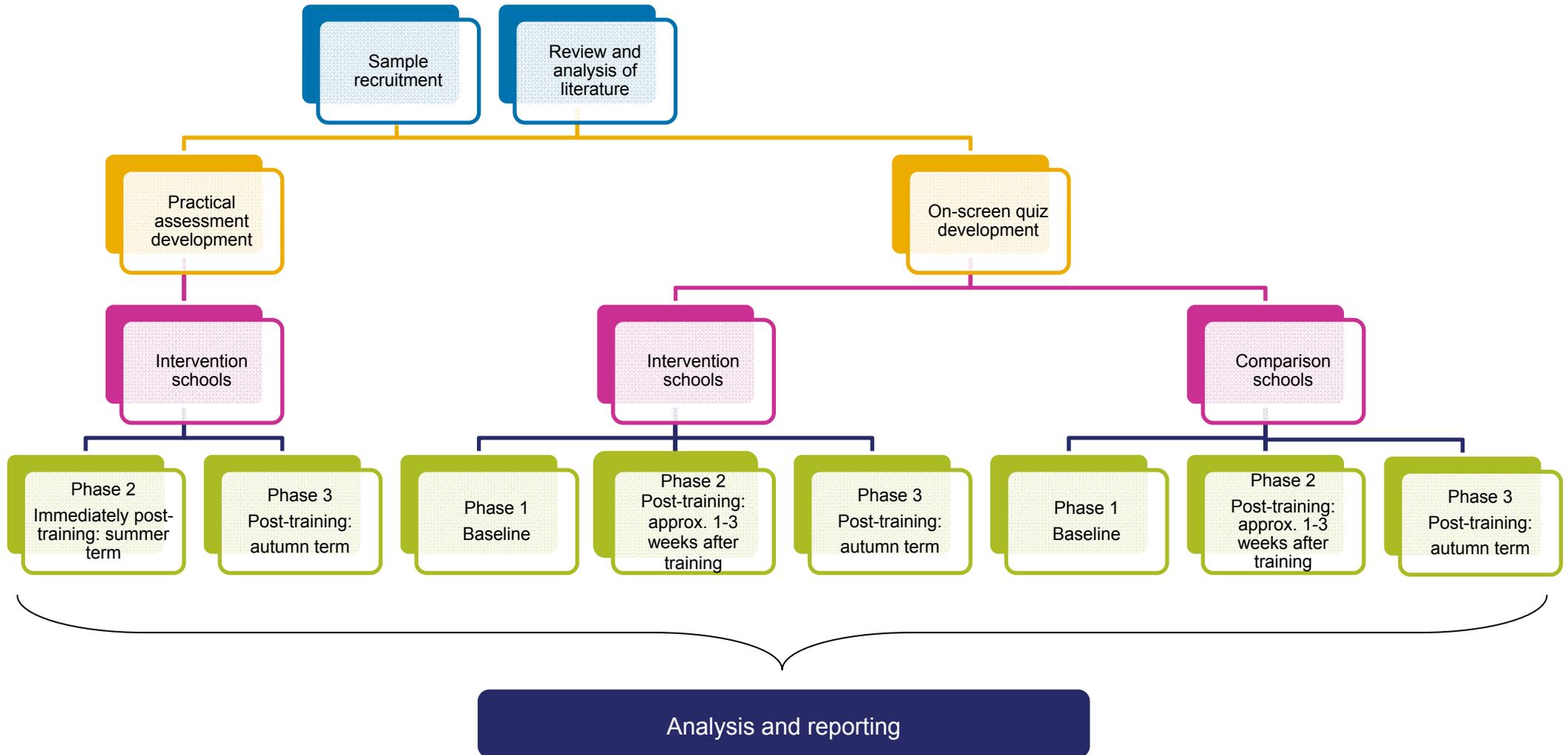
## 2.2 Review and analysis of relevant accident data and literature relating to cycle training

The purpose of the review of relevant collision and injury data and literature relating to cycle training was to consider the most common risks children face when cycling on the road. This was used to inform the design of the on-screen quiz and practical assessments for testing children's hazard perception and risk mitigation ability.

The review first considered common risks faced by children riding on the road and the factors which affect their ability to develop skills of hazard perception. It then identified common on-road conflicts including those involving child cyclists; factors contributing to these conflicts; and the effectiveness of cycle training in reducing risk for children.

The review is provided in full in Appendix A.

The most relevant findings from the review were:

- In the UK, the risk of someone who cycles being killed or seriously injured is reported to be highest for young cyclists aged 10-15 years.

- Police-reported injury road collisions data indicates that over four-fifths of killed or seriously injured (KSI) cycle collisions are as a result of an impact with another vehicle - with cyclist 'crossing or entering road into path of vehicle' being a frequent cause of KSI for child cyclists.

- The two main contributory factors assigned to child cyclists involved in collisions were that the child 'failed to look properly' and 'entered the road from the pavement'.

- Depending on their age, children can have serious knowledge, perceptual and cognitive limitations in relation to roads. They can be unpredictable, do not have a good appreciation of road hazards and are generally unfamiliar with road rules.

- By the age of 10, children can achieve basic cycling competence with appropriate training for riding on quiet two-lane roads, negotiating parked cars and simple junctions.

- To date, most evaluations of cycle training have either focused on cycle training in the UK before Bikeability was introduced or on cycle training delivered in other countries. Of these studies outside the UK, there were two evaluations of cycle training programmes for which effect sizes of between 1.3 and 2.1 were reported (though it is noted that neither programme was directly comparable with Bikeability Level 2).

The findings from the review were used alongside the outcomes of the National Standard for Cycle Training Level 2 to inform the content and types of questions required for the on-screen quiz development.

## 2.3 On-screen quiz development

The on-screen quiz was designed to enable the assessment of children's responses to a wide variety of situations in which hazards may occur on typical Bikeability Level 2 single-lane roads and junctions with varying degrees of complexity and traffic.

A bank of questions was developed from which two versions of the on-screen quiz were created. Each version of the on-screen quiz was designed to take approximately 30-40 minutes to complete and comprised 35 hazard perception and risk mitigation questions together with some additional questions collecting background pupil data (e.g. age, gender, cycling frequency, on-road cycling experience, cycling confidence). Each version contained 15 common questions to enable the performance of each question and of each pupil to be compared statistically across both versions using an item response theory (IRT) model (please refer to Appendix E.3 for further explanation).

The quizzes were based on the cycling journeys of three children – Sam, Eva and Ben. This allowed for the quiz to be contextualised and for features of a journey to be modelled (i.e. preparing for a journey, getting started, riding on the road, dealing with junctions and ending a journey). It also ensured that 'real-life' cycling situations could be presented. The questions were all closed response format and were designed to be appealing to pupils through the use of real-life photographs and film clips of the children on their journeys.

Further details of the on-screen quiz development are provided in Appendix B.

## 2.4 Practical assessment development

For the purpose of this research, on-road cycling behaviour related to hazard perception and risk mitigation was assessed in an observational session at two time points – one to three weeks after training (phase 2) was completed in the summer term and again two to three months later (phase 3), the following September. In order to remove any 'bias' in the data collection, pupils were assessed by qualified, experienced and fully briefed National Standard instructors (NSIQs) who had not delivered their training.

In order to minimise the potential risk to children participating in the study, only children who had been trained and had passed Bikeability at Level 2 were allowed to participate in the practical assessments. Parents / guardians also had to provide informed consent. The NSIs delivering the practical assessments applied standard Bikeability risk assessment and control measures covering equipment (bicycles, helmets), trainee preparedness, environment (routes to and sites used for assessment) and dynamic risk assessment (changing conditions encountered during assessment). In addition, assessors were instructed to use junctions that would typically be used for the first day of a Level 2 training course for the practical assessment.

The practical assessment took place in the period after training finished before the end of the summer term – in most cases this was within a two-week period, but this varied from school to school. The second practical assessment took place during the first half of the autumn term.

Prior to the first practical assessment session, NSI assessors were provided with training on how to complete the assessments. During the assessments, they recorded results on a 'pupil record sheet' and this data was captured and matched to the pupils' outcomes on the on-screen quiz.

It is acknowledged that there may be a perceived bias in using NSIs to assess whether Bikeability is effective in improving children's hazard perception and risk mitigation. However, it was considered to be both unethical and impractical for these assessments to be done without extensive experience of assessing children's on-road cycling abilities.

In order to minimise bias, all practical assessments were conducted by NSIs from a neighbouring area or who had not previously trained or assessed any of the children in the assessment group.

More importantly, the practical assessment provides only one piece of evidence about children's hazard perception and risk mitigation abilities. The main purpose of the practical assessment was to check the validity of the on-screen quiz.

Further details of the practical assessment development are provided in Appendix C.

## 2.5 Sample recruitment

In order to assess the immediate and longer-term impact of Bikeability training on children's hazard perception and risk mitigation ability, it was important to test children soon after the completion of training and also some months later. Due to the desirability of carrying out the practical assessment in the warmer months, the ideal period for carrying out the hazard perception and risk mitigation research was during the summer term, repeating the assessment a minimum of two months later, in September 2014.

The project involved pupils who were in Year 5 (Y5) in the summer term and who moved into Year 6 (Y6) in September 2014 so that they would be of a suitable age and also easy to track within the same school.

Two samples of pupils were identified – those who would be trained whilst in Year 5 (the 'intervention' group) and those who would be trained when in Year 6 (the 'comparison' group). Eleven local authorities/London boroughs were approached and Bikeability delivery schemes were invited to participate in the study. Of these, six schemes agreed to take part (all London boroughs are treated as one area for the purposes of reporting) and provided names of schools that met the criteria for selection (i.e. having pupils that would be trained in Y6 for the comparison group, or having Y5 pupils that were due to be trained in the summer term for the intervention group). A total of 335 schools were approached and 27 agreed to participate.

Recruitment proved problematic within the tight project timescales and due to the complex logistical requirements, so a top-up sample of schools was drawn requesting participation at one time point, in order to collect data for the on-screen quiz questions. This resulted in an additional 16 schools agreeing to participate. Having recruited sufficient schools in advance, actual participation was much lower than expected. Due to the level of drop-out between baseline and post-training assessment, ten additional schools were also approached for the final assessment period to further boost data on the on-screen quiz questions. A total of 29 schools took part in at least one phase of the research.

A summary of the numbers of schools recruited and completing the assessments at each of the three time points are shown in Table 2.1.

Further details of the recruitment process are provided in Appendix D.

**Table 2.1** **A summary of school recruitment and assessment completion**

| | Area | Summer term (Y5 pupils) | | | | | Autumn term (Y6 pupils) | | |
| | | No. of schools invited | No. agreeing to participate | No. of schools participating | | | No. of schools invited | Number of schools participating | |
| | | | | On-screen phase 1 | On-screen phase 2 | Practical phase 2 | | On-screen phase 3 | Practical phase 3 |
|---|---|---|---|---|---|---|---|---|---|
| **Intervention** | A | 11 | 3 | 3 | 2 | 2 | 2 | 2 | 2 |
| | B | 10 | 2 | 2 | 1 | 2 | 2 | 2 | 2 |
| | C | 9 | 2 | 2 | 1 | 2 | 2 | 2 | 2 |
| | D | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | E | 21 | 2 | 2 | 1 | n/a | 2 | 1 | n/a |
| **Comparison** | A | 55 | 10 | 4 | 1 | n/a | 4 | 2 | n/a |
| | B | 10 | 3 | 1 | 1 | n/a | 1 | 0 | n/a |
| | C | 6 | 2 | 1 | 0 | n/a | 1 | 1 | n/a |
| | D | 6 | 1 | 0 | 0 | n/a | 0 | 0 | n/a |
| | E | 8 | 2 | 2 | 1 | n/a | 2 | 1 | n/a |
| **Top up sample 1** | Random, national | 200 | 16 | 10 | - | - | - | - | - |
| **Top up sample 2** | Areas A-E | - | - | - | - | - | 10 | 3 | - |
| **Total** | | **335** | **43** | **27** | **8** | **6** | **26** | **14** | **6** |

## 2.6 A summary of assessment completion by group and phase

**Intervention group**

**Phase 1 (P1)**
- **Summer term 2014: Baseline**
- On-screen quiz

**Phase 2 (P2)**
- **Summer term 2014: Post-training**
- On-screen quiz
- Practical assessment

**Phase 3 (P3)**
- **Autumn term 2014: Post-training**
- On-screen quiz
- Practical assessment

**Comparison group**

**Phase 1 (P1)**
- **Summer term 2014: Baseline**
- On-screen quiz

**Phase 2 (P2)**
- **Summer term 2014**
- On-screen quiz

**Phase 3 (P3)**
- **Autumn term 2014**
- On-screen quiz
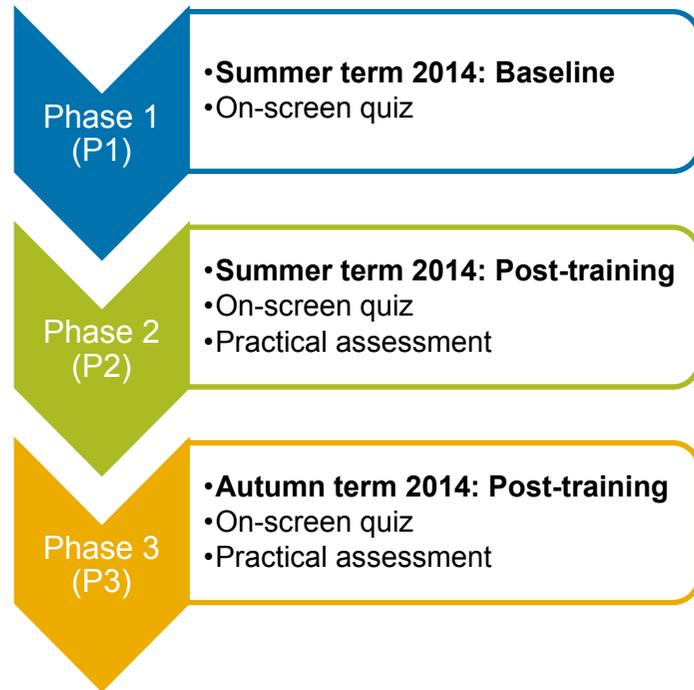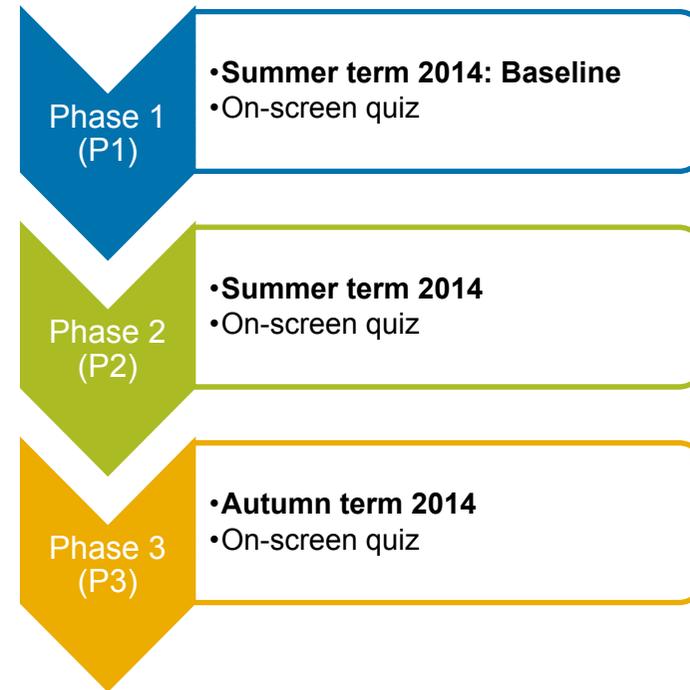
**On-screen quiz**



**Practical assessment**

## 2.7 Analysis

The objective of the analysis was to measure the association between the training and a number of outcome measures, by comparing the average outcomes in the intervention group with the average outcomes in the comparison group. The primary outcome measure of interest is from the on-screen quiz. We also looked at a number of attitudinal measures based on responses to the background section of the quiz.

Multilevel regression modelling was used to estimate the average difference between the intervention and comparison groups. Multilevel modelling is a statistical method that takes account of the fact that pupils are nested within schools and that the ability of pupils in the same school will therefore be correlated. Including the outcome variable measured at phase 1 as an explanatory variable in a multilevel regression model also takes account of each pupil's underlying ability, increasing the power of the analysis.

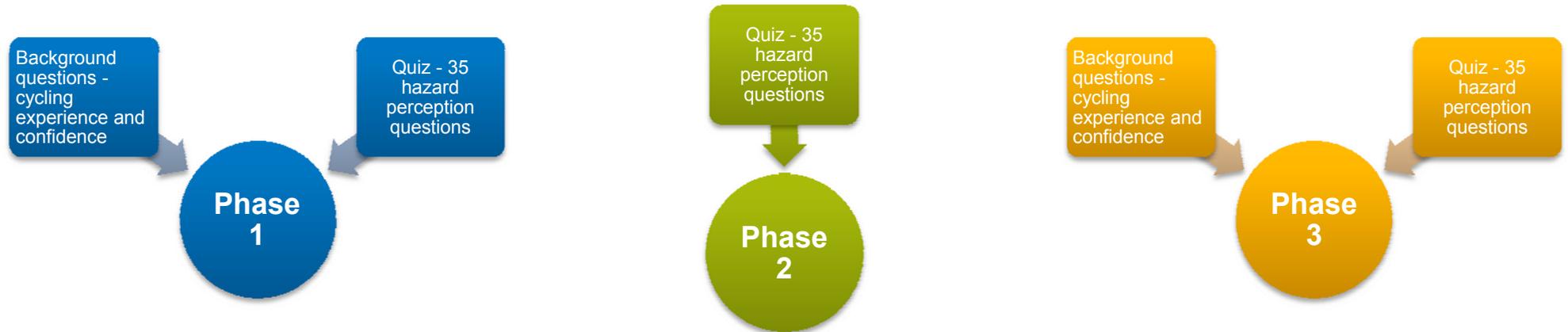However, the average differences between pupils in the two groups should be interpreted cautiously because they may reflect systematic differences between the two groups that affect the outcomes, but that we have not measured. A randomised controlled trial, in which schools or pupils are randomly allocated to an intervention (in this case, Bikeability training) or control/comparison group, is recognised as the most robust way to determine causation. However, it was not possible to employ this methodology for this study because we had no control over when the training was taking place.

A number of factors may increase the confidence that we have about whether the differences in our analysis demonstrate the impact of the training. The time between the phase 1 (baseline) and phase 2 (one to three weeks after training) was very short, which reduced the likelihood that the observed differences are likely to be due to other factors such as age. Also, because pupils' outcome measures at baseline were included as an explanatory variable in the statistical modelling, we took account of any underlying differences between the two sets of pupils relating to their pre-existing abilities and attitudes.

# 3 Outcomes

## Key Findings

- The average 'hazard perception and risk mitigation ability' of pupils in the intervention group was much higher, after training, than in the comparison group who had received no training.

- The association between training and increased 'hazard perception and risk mitigation ability' was undiminished at phase 3, suggesting the association with training was sustained.

- The size of the association between training and hazard perception is very large, with an effect size of 1.6.

- The on-screen quiz functioned appropriately with a reliability measure of 0.76 (Cronbach's Alpha) indicating that it discriminates well between pupils who achieve higher and lower hazard perception and risk mitigation scores.

- The analysis of attitudes shows a statistically significant association between the training and increased cycling confidence.

- There is no association between training and an increase in frequency of cycling.

- In the on-screen quiz, 'observation' was the highest scoring domain across all phases.

- In the practical assessments, in both phases (2 and 3) 'observation' was the highest scoring domain whilst 'communication' was the lowest scoring.

- The correlation between the practical assessment and the on-screen quiz was positive and statistically significant, which confirms that the practical assessment does validate the on-screen quiz.

## 3.1 Summary of participation

Table 3.1 summarises the data collected at each time point whilst Table 3.2 provides information about the numbers of pupils completing each combination of assessments.

**This document is a confidential draft**

**Table 3.1    Summary of participation – all pupils**

| Assessment | Number of pupils completing each assessment | | | | |
|---|---|---|---|---|---|
| | Intervention schools | Intervention schools (London) | Comparison schools | Comparison schools (top-up) | Total |
| Phase 1 quiz | 118 | 20 | 220 | 159 | 517 |
| Phase 2 quiz | 72 | 7 | 91 | - | 170 |
| Phase 2 practical assessment | 75 | n/a | n/a | n/a | 75 |
| Phase 3 quiz | 117 | 12 | 70 | 118 | 317 |
| Phase 3 practical assessment | 64 | n/a | n/a | n/a | 64 |

**Table 3.2    Summary of assessment combinations completed – matched pupils**

| Assessment combination | Numbers of pupils completing each combination | | | | |
|---|---|---|---|---|---|
| | Intervention schools | Intervention schools (London) | Comparison schools | Comparison schools (top-up) | Total |
| Phase 1 quiz | 118 | 20 | 220 | 159 | 517 |
| Phase 1 quiz + Phase 2 quiz | 66 | 0 | 76 | 0 | 142 |
| Phase 1 quiz + Phase 2 practical | 61 | n/a | n/a | n/a | 61 |
| Phase 1 quiz + Phase 2 quiz + Phase 2 practical | 43 | n/a | n/a | n/a | 43 |
| Phase 1 quiz + Phase 2 quiz + Phase 3 quiz | 60 | 0 | 27 | 0 | 87 |
| Phase 1 quiz + Phase 2 quiz + Phase 2 practical + Phase 3 quiz + Phase 3 practical | 37 | n/a | n/a | n/a | 37 |
| Phase 1 quiz + Phase 3 quiz | 95 | 6 | 53 | 0 | 154 |
| Phase 3 quiz | 0 | 0 | 0 | 118 | 118 |

## 3.2 On-screen quiz outcomes

Pupils took one of two versions of the quiz each of which contained 35 questions. There were 15 questions that were the same in both versions of the quiz and 20 questions that were unique to each version.

Initial analysis of the responses to the quiz was used to establish the functioning of the individual quiz questions. This was done to ensure that the questions performed as required – i.e. to check that the quiz included questions with a range of difficulty and discriminated between those with higher and lower 'hazard perception and risk mitigation ability'. Initial analysis revealed that of the 55 unique questions across the two quizzes, 14 did not function appropriately – for example, they were too easy or did not discriminate between the different levels of performance. These questions were removed from any further analysis. Further details of the functioning of each question are provided in Appendix E.1.

Following the initial analysis, each pupil's responses to the questions in the on-screen quiz were used to derive a measure of their 'hazard perception and risk mitigation ability'. Where there are two or more versions of a quiz, a simple raw score (i.e. the percentage of correct answers) does not take into account that one version might be more or less difficult depending on the difficulty of the unique questions presented. Therefore, we used item response theory (IRT) to estimate a measure of pupils' 'hazard perception and risk mitigation ability' that took account of the difficulties of each question and measured ability on the same scale for each pupil (see Appendix E.3 for more information about IRT). The 'hazard perception and risk mitigation ability' scale was designed to have a mean of 100 and standard deviation of 20 at phase 1, which meant that around 70 per cent of pupils had scores between 80 and 120 and around 95 per cent had scores between 60 and 140.

Table 3.3 displays the mean and standard deviation of the ability measure by phase and group. The average ability in the intervention and comparison groups is similar, which gives added confidence that there are no important underlying differences between the two groups.

**Table 3.3    Hazard perception and risk mitigation ability summary statistics**

| Quiz phase | Group | Mean ability | Standard deviation | Number of pupils |
|---|---|---|---|---|
| Phase 1 quiz (baseline) | Intervention | 102.6 | 16.7 | *138* |
|  | Comparison | 99.0 | 21.0 | *379* |
|  | All | 100.0 | 20.0 | *517* |
| Phase 2 quiz (1-3 weeks after training) | Intervention | 129.7 | 26.9 | *79* |
|  | Comparison | 104.7 | 25.0 | *91* |
|  | All | 116.3 | 28.7 | *170* |
| Phase 3 quiz (2-3 months after training) | Intervention | 130.5 | 22.2 | *129* |
|  | Comparison | 104.0 | 17.3 | *188* |
|  | All | 114.8 | 23.4 | *317* |

Although the intervention group had a slightly higher 'hazard perception and risk mitigation ability' score at the baseline than the comparison group, this difference was not statistically significant.

The data shows that the average 'hazard perception and risk mitigation ability' of pupils in the intervention group was much higher after training than in the comparison group. The differences between the mean ability scores at phase 1 and phase 2/3 are statistically significant. Further discussion of these results is provided in section 3.4 below.

The statistics shown in Table 3.3 describe the results of all the pupils that took the quiz at each time point, including many pupils that only took one quiz. The results in section 3.4 focus on those pupils that took more than one quiz, allowing us to compare the progress they made over time and look more closely at the impact of the training.

In addition to looking at how the quiz functioned as a whole, the quiz questions were developed and analysed in four domains – observation, communication, road position and priorities – in an attempt to explore the supplementary research questions:

- Do trained children select safer cycling strategies – i.e. observation, communication, road position, priorities?

  o  Observation - are they more aware of other road users?

  o  Communication - are they better at knowing when to signal their intentions (and when not), and how to communicate to best effect?

  o  Road position - do they select optimal road positions for different phases of their journey?

  o  Priorities - do they have a better understanding of their priorities and those of other road users?

The results of the quizzes, by the four domains, are provided in Table 3.4 below. It should be noted that some of the questions were not allocated to a domain because they were more associated with Level 1 training (for example, questions about getting ready to ride) and some were allocated to more than one domain where more than one skill was being assessed. Further detail is provided in Appendix E.1.

**Table 3.4**     **'Hazard perception and risk mitigation ability' statistics by domain**

| Phase | Domain | Number of quiz questions | Percentage of answers correct within the domain | |
|---|---|---|---|---|
| | | | Intervention group | Comparison group |
| Phase 1 (baseline) | Observation | 16 | 41% | 38% |
| | Communication | 7 | 35% | 30% |
| | Road position | 10 | 21% | 21% |
| | Priorities | 8 | 35% | 33% |
| | *Number of pupils* | | *138* | *379* |
| Phase 2 (1-3 weeks after training) | Observation | 16 | 55% | 40% |
| | Communication | 7 | 44% | 34% |
| | Road position | 10 | 46% | 25% |
| | Priorities | 8 | 51% | 33% |
| | *Number of pupils* | | *79* | *91* |
| Phase 3 (2-3 months after training) | Observation | 16 | 58% | 43% |
| | Communication | 7 | 44% | 33% |
| | Road position | 10 | 43% | 19% |
| | Priorities | 8 | 51% | 38% |
| | *Number of pupils* | | *129* | *188* |

These results show that, across all three phases, children score most highly on questions about 'observation'. In phases 1 and 3, they score most poorly on questions about 'road position'. However, it is also the case that the biggest increase in score for the intervention group was in 'road position'. It is noteworthy that the numbers of questions in each domain is variable and in some cases quite small, which means that these results should be treated with some degree of caution. These results will be further examined in sections 3.4 and 3.6 - 3.7 below.

## 3.3 Attitude outcomes

As well as measuring the difference in 'hazard perception and risk mitigation ability' between the intervention and comparison groups, we looked at the difference in cycling confidence, cycling enjoyment and frequency of cycling to see whether training was associated with an increase in any of these attitude measures. The measures were taken from the background questionnaires at phase 1 and phase 3, and were defined as:

- **Confidence**

    How confident do you feel about riding a bicycle on the road?

    from 1 = 'not at all confident' to 4 = 'very confident'

- **Enjoyment**

    Do you enjoy cycling?

    from 1 = 'never' to 5 = 'always'

- **Frequency**

    How often do you ride a bicycle?

    from 1 = 'never' to 6 = 'every day'.

The outcomes of this analysis are discussed in section 3.4 below.


## 3.4 Impact evaluation

We compared average 'hazard perception and risk mitigation ability' scores in the intervention group and the comparison group in order to measure the association between the training and hazard perception. The comparison took account of each pupil's underlying ability by including the ability at phase 1 as an explanatory variable in a multilevel regression model. A graphical representation of the multilevel regression model is shown in Figure 3.1, where ability at phase 1 is shown along the horizontal axis and ability at phase 2 is shown on the vertical axis. Pink dots show the data from the intervention group pupils and blue dots show the data from the comparison group pupils. Regression modelling fits a *line of best fit* through the data and simultaneously measures the difference in ability at phase 2 that is associated with the training.

**Figure 3.1** **'Hazard perception and risk mitigation ability' at phases 1 and 2, in the intervention and comparison groups**

Table 3.5 displays a summary of the impact evaluation findings, showing the association between pupils' involvement in training and average 'hazard perception and risk mitigation ability'. The analysis shows that the average ability of pupils in the intervention group was much higher after training than in the comparison group. Further, the association between training and increased ability is undiminished at phase 3, suggesting the association with training is sustained.

**Table 3.5    Summary of impact evaluation findings – 'hazard perception and risk mitigation ability'**

|  | Difference in score* | Statistically significant?** | Effect size | Number of pupils (intervention / comparison) |
|---|---|---|---|---|
| **'Hazard perception and risk mitigation ability'** | | | | |
| Phase 2 | 28.3 | Yes | 1.58 | *66 / 76* |
| Phase 3 | 28.7 | Yes | 1.60 | *101 / 53* |

* The difference in score is calculated by subtracting the mean score of the comparison group from the mean score of the intervention group, while taking account of each pupil's score at phase 1.
** Statistical significance is at the 95% level.

The analysis provides compelling evidence of a relationship between training and increased 'hazard perception and risk mitigation ability'.

Statistical testing suggests the results are unlikely to be due to chance. Due to the average ability scores in the two groups being similar at phase 1, and any differences having been taken into account in the model, it is also unlikely to be due to pre-existing differences between the two groups.

The size of the association between training and hazard perception is very large, with an effect size of 1.6. However, that size of effect is consistent with other similar studies (see Appendix A):

- McLaughlin and Glang (2010) found an effect size of the 'Bike Smart' cycle training programme of 2.05 on hazard discrimination and 1.42 on safety rules

- Ducheyne et al. (2013) found an effect size of 1.30 on a measure of practical cycling skill.

The effect size is much larger than most educational interventions in the UK (for example, see the Sutton Trust EEF Teaching and Learning Toolkit), but that is to be expected because knowledge of hazard perception and risk mitigation among the pupils at baseline is low and the skills being assessed are very well aligned with the training.

Table 3.6 displays a summary of the impact findings relating specifically to the pupils' scores by domain on the on-screen quiz. The analysis shows that the training was

associated with gains in all areas, and that the largest gains were in the domains of 'road position' and 'priorities'.

**Table 3.6     Summary of impact evaluation findings – 'hazard perception and risk mitigation ability' by domain**

| | Difference in score (pp)* | Statistically significant?** | Effect size | Number of pupils (intervention / comparison) |
|---|---|---|---|---|
| **'Hazard perception and risk mitigation ability' domains** | | | | |
| **Phase 2** | | | | |
| Observation | 17 | Yes | 1.05 | *66 / 76* |
| Communication | 13 | Yes | 0.77 | *66 / 76* |
| Road Position | 22 | Yes | 1.39 | *66 / 76* |
| Priorities | 24 | Yes | 1.21 | *66 / 76* |
| **Phase 3** | | | | |
| Observation | 17 | Yes | 1.07 | *101 / 53* |
| Communication | 13 | Yes | 0.75 | *101 / 53* |
| Road Position | 24 | Yes | 1.54 | *101 / 53* |
| Priorities | 20 | Yes | 1.02 | *101 / 53* |

\* The difference in score is calculated by subtracting the mean score of the comparison group from the mean score of the intervention group, while taking account of each pupil's score at phase 1. Domain scores are the percentage of correct answers, so the differences are expressed in percentage point (pp) terms, i.e. the difference between 40% and 60% is 20 percentage points.
\*\* Statistical significance is at the 95% level.

Although, as shown in Table 3.4, children's scores are relatively low for 'road position' and 'priorities', the fact that there is a larger percentage point gain as a result of training is encouraging.

Analysis of the results by domain is helpful for identifying areas of the training which currently benefit children to a greater or lesser extent. Scores achieved in 'communication' are below half marks at all phases and whilst children make significant improvements in the domains of 'road position' and 'priorities', the scores achieved in these domains are still relatively low. This indicates that it would be beneficial for these areas to receive more attention from Bikeability providers.

Table 3.7 displays a summary of the impact findings relating to attitude measures from the background questions. The analysis shows a statistically significant association between the training and increased confidence cycling on the road.

**Table 3.7    Summary of impact evaluation findings – attitudes**

| | Difference in score* | Statistically significant?** | Effect size | Number of pupils (intervention / comparison) |
|---|---|---|---|---|
| **Attitude measures** | | | | |
| Confidence | 0.47 | Yes | 0.53 | *99 / 53* |
| Enjoyment | 0.19 | No | 0.20 | *94 / 47* |
| Frequency | -0.02 | No | -0.02 | *98 / 51* |

* The difference in score is calculated by subtracting the mean score of the comparison group from the mean score of the intervention group, while taking account of each pupil's score at phase 1.
** Statistical significance is at the 95% level.

Whilst there is a small association with increased enjoyment of cycling, it is not statistically significant, which means we cannot be confident that the difference is not due to chance. There appears to be no relationship between training and increased frequency of cycling.

## 3.5 Practical assessment outcomes

In the practical assessment pupils were assessed on their ability to demonstrate competence, confidence and consistency in the four domains of: observation, communication, road position and priorities. Pupils were assessed by two assessors, one assessor per drill. They were given a score between zero and three for each domain depending on the extent to which the assessors observed their skills of hazard perception and risk mitigation in the four domains. The following scale was used:

0 = never observed; 1 = rarely observed; 2 = mostly observed; 3 = always observed.

For example, in order to perceive hazards and mitigate risks associated with passing a parked car, pupils' observation (looking behind before passing the car) and road position (passing the car with enough room to clear an open door) were each scored. Each drill was performed twice and children had ten minutes in which to complete the whole assessment, providing sufficient opportunities for assessors to assess pupils' performance in each domain area.

Only pupils who had passed the Bikeability Level 2 training were assessed on their practical skills in the intervention group. They were assessed at phase 2 (one to three weeks after training was completed) and phase 3 (two to three months after training was completed). Table 3.8 shows the mean score, standard deviation and range of scores for each domain of the practical assessment, as well as the overall score.

**Table 3.8**     **Summary of practical assessment scores**

| Phase | Domain | Mean score | Standard deviation | Range | Number of pupils |
|---|---|---|---|---|---|
| Phase 2 (1-3 weeks after training) | Observation | 4.1 | 1.4 | 0 – 6 | 75 |
| | Communication | 3.3 | 1.3 | 0 – 6 | 75 |
| | Road position | 3.7 | 1.4 | 1 – 6 | 75 |
| | Priorities | 3.9 | 1.7 | 0 – 6 | 75 |
| | **Total** | **15.0** | **4.8** | **3 – 24** | **75** |
| Phase 3 (2-3 months after training) | Observation | 3.8 | 1.5 | 0 – 6 | 64 |
| | Communication | 2.8 | 1.4 | 0 – 6 | 64 |
| | Road position | 3.0 | 1.4 | 0 – 5 | 64 |
| | Priorities | 3.7 | 1.2 | 1 – 6 | 64 |
| | **Total** | **13.2** | **4.1** | **4 – 22** | **64** |

'Observation' is the highest scoring domain at both time points whilst 'communication' is the lowest scoring. It is interesting that 'road position' shows the greatest decrease between phase 2 and phase 3, which suggests that pupils are least likely to maintain this skill over time.

There were 64 pupils who were assessed twice (at phases 2 and 3). The mean scores achieved by these pupils were 14.8 at phase 2 and 13.2 at phase 3 and analysis shows that the decrease in the mean score for this group of pupils (1.6 marks) was statistically significant.

The wide range of scores achieved on the practical assessment, particularly for phase 2 which were based on assessments carried out within one to three weeks of training taking place, is noteworthy.  It is reasonable to expect that children who had just passed the training should score more highly on this assessment and that there should be some consistency in their scores, i.e. that those children who had just passed the training should be achieving three marks in each domain by demonstrating competence, confidence and consistency in the skill.

It is possible that this variability may, in part, be explained by inter-assessor variability. However, analysis revealed that 75 per cent of pupils' two ratings were within one point (out of 12) of each other and 95 per cent within three points. We also found that Cronbach's alpha (a measure of reliability) of the two measures was high (0.85), suggesting the reliability of the measures was good.

The variability in scores may be indicative of some inconsistency in how different Bikeability schemes and instructors train and assess pupils at Level 2.

## 3.6 Comparing the on-screen quiz and practical assessment outcomes

When considering the mean score on the assessments as a whole, it is noteworthy that whilst the on-screen quiz score remains fairly static between phase 2 and 3 (i.e. there is no indication of a decrease in the acquired knowledge relating to hazard perception and risk mitigation), there is a statistically significant decrease in the mean score obtained on the practical assessment two to three months after the training has taken place. This suggests that the ability to put the knowledge into practice declines over time. This may be related to the fact that the frequency of children's cycling does not increase following training.

It is interesting to compare the outcomes of the on-screen quiz and practical assessment when split by domain. 'Observation' is the highest scoring domain in both forms of assessment across both phases whilst 'communication' is the lowest scoring domain in both forms of assessment at phase 2 and in the practical assessment at phase 3. Although the scores achieved on 'priorities' in both forms of assessment at phases 2 and 3 remain fairly consistent, it is interesting that scores in the 'road position' domain decrease between phases 2 and 3 in both forms, though in a more pronounced way in the practical assessment. This might suggest that whilst children maintain some knowledge of road positioning, they are less likely to demonstrate it in practice.

## 3.7 Correlation between on-screen and practical assessment scores

The practical assessment was used to check the validity of the on-screen quiz. We can assess whether the on-screen quiz and the practical assessment are likely to be measuring the same underlying construct by looking at the correlation between pupils' on-screen quiz and practical assessment scores.

The correlation between the two scores was 0.40 at phase 2 and 0.35 at phase 3. Both correlation coefficients were positive and statistically significant, which suggests the practical assessment does validate the on-screen quiz.

Figure 3.2 presents a graphical representation of the correlation between the practical assessment score and the on-screen quiz score at phase 2, where the dots show the data for each pupil and the line shows a *line of best fit*.

**Figure 3.2    Correlation between on-screen quiz score and practical assessment score at phase 2**
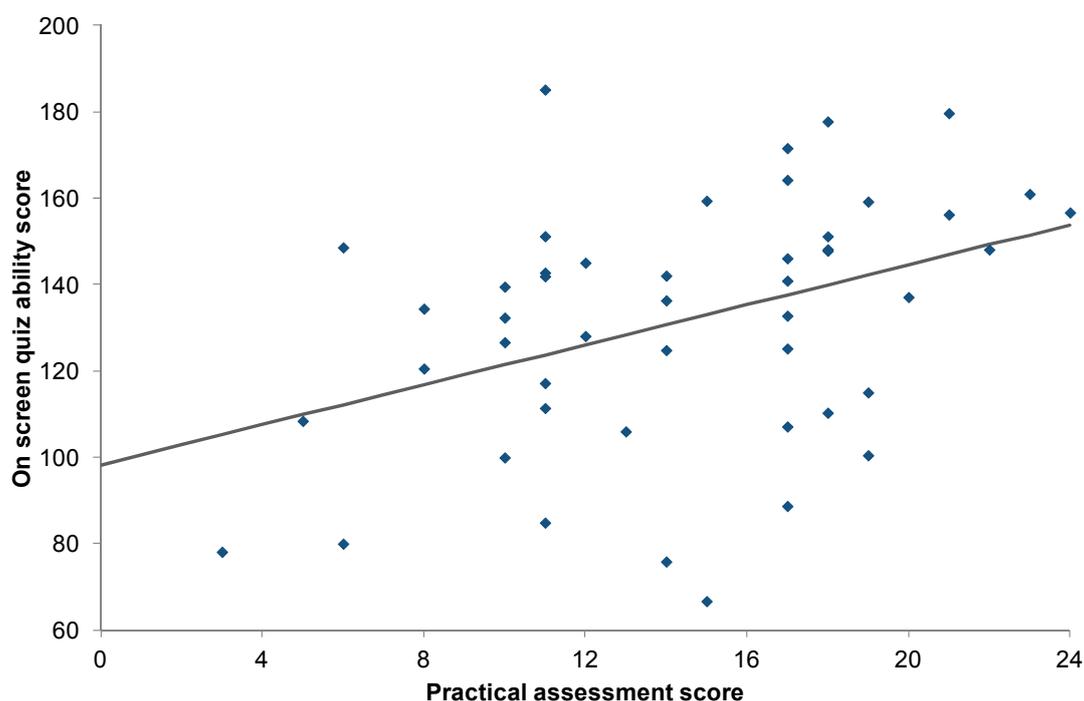


Table 3.9 summarises the correlations between the on-screen quiz and the practical assessment scores in the four domains of observation, communication, road position and priorities.

**Table 3.9    Correlation between on-screen quiz and practical assessment scores**

| Domain | Phase 2 | | Phase 3 | |
|---|---|---|---|---|
| | Correlation | Significant? | Correlation | Significant? |
| Observation | 0.48 | Yes | 0.28 | Yes |
| Communication | 0.28 | No | 0.29 | Yes |
| Road position | 0.22 | No | 0.24 | No |
| Priorities | 0.40 | Yes | 0.02 | No |
| Overall | 0.40 | Yes | 0.35 | Yes |
| *Number of pupils* | *48* | | *60* | |

While the overall measures seem to be measuring the same construct, there is some doubt as to whether the individual domains, particularly road position, are well aligned to each other.

As discussed in section 3.2, the number of questions in each of the domains in the on-screen quiz was variable and this is likely to have impacted on the correlation

analysis due to the small number of pupils taking each of the components (i.e. children who took both the on-screen quiz and practical assessment) and the small number of questions used for comparison (i.e. only four score points/marks on the practical assessment and a maximum of seven marks on one domain in the on-screen quiz).

It is worth bearing in mind that the number of pupils is small, making the precision of the estimates low and thus results should be treated with caution.

# 4 Conclusions and recommendations

## 4.1 Conclusions

Road accident data indicates that it is young cyclists aged 10-15 who are most at risk of being killed or seriously injured due to a collision with another vehicle. Furthermore, collisions resulting from cyclists 'crossing or entering road into the path of another vehicle' are particularly frequent for child cyclists.

Although, as Turner *et al.* (2009) tell us, children can have serious knowledge, perceptual and cognitive limitations in relation to roads and can be unpredictable because they do not have a good appreciation of road hazards, by the age of ten, they can achieve basic cycling competence with appropriate training for riding on quiet two-lane roads, negotiating parked cars and simple junctions.

Bikeability aims to give children the skills and confidence they need to cycle on today's roads, and so encourage more people to cycle more often with less risk. In order to ascertain whether Bikeability Level 2 training improves children's hazard perception, NFER developed a bank of questions which were successfully used to establish a 'hazard perception and risk mitigation ability' measure.

A total of 41 questions functioned appropriately with a reliability measure of 0.76 (Cronbach's alpha), demonstrating that the on-screen quiz is an effective and reliable tool for measuring pupils' hazard perception and risk mitigation ability.

The results of the on-screen quiz indicated that pupils who had participated in Bikeability Level 2 training scored significantly higher than pupils of the same age who had not received any training. Analysis shows that the training has a large effect size of 1.6 and that this effect is sustained over a period of at least two months following training.

There was a significant decrease in the mean scores achieved on the practical assessment between phase 2 and phase 3. This suggest that whilst trained children achieved higher scores for the on-screen quiz and sustained this over a period of time, the ability to put that knowledge into practice can decline over time if the skills are not practised.

Pupils' 'hazard perception and risk mitigation ability', as measured by the on-screen quiz, was positively correlated with their practical assessment scores, indicating that the on-screen quiz and practical assessment are measuring a similar construct.

In addition to measuring the effects of the training on pupils' hazard perception and risk mitigation ability, NFER also monitored changes in pupils' attitudes towards cycling. Analysis shows that there is a statistically significant positive association between the training and increased cycling confidence. However, children who participated in training did not cycle more often after training.

## 4.2 Recommendations

We recommend that further use is made of the bespoke, validated on-screen quiz questions. A finalised version of the quiz could be constructed using some or all of the 41 questions which proved to function appropriately. The quiz could then be used in the following ways:

| Use | Purpose |
|---|---|
| Gather baseline and post-training data from a larger sample of children. | This will provide detail about the functioning of the quiz in a finalised format and build on the individual question functioning gathered during this research. This information can be used to inform future use. |
| | This would also enable more detailed analysis of the impact of training on different sub-groups of children (e.g. by age, gender, ethnicity). |
| Use the finalised quiz at baseline and post-training with children taking the Bikeability Level 2 training. | This will allow monitoring of the effectiveness of the Bikeability training and identify any particular areas which may need development. For example, if it reveals that pupils do not score well on questions about 'road position', it may be that more attention is required in this area of training. |
| | It will provide a means of evaluating/monitoring the consistency and quality of the training given by different providers across the country. |
| | It will provide a means of evaluating/monitoring and moderating the different training delivery models. |
| Use the finalised quiz at baseline, immediately post-training and again after a longer period of time. | This will provide further information about how sustained the observed effects of the training are over time – i.e. we have established that children maintain their hazard perception and risk mitigation skills for at least two months, but further analysis could be done to see if the training has even greater longevity. |
| | This may also be useful in establishing if there are any areas of the training for which follow–up or refresher training may be usefully implemented. |

# References

Bikeability (n.d.) *Bikeability Delivery Guide: Delivery Guidance for Instructors and Training Providers*. London: Department for Transport [online]. Available: http://bikeability.dft.gov.uk/wp-content/uploads/Bikeability_Delivery_Guide.pdf [4 December, 2014].

Bikeability (2014). *What is Bikeability?* [online]. Available: https://bikeability.dft.gov.uk/what-is-bikeability/ [5 December, 2014].

Ducheyne, F., De Bourdeaudhuij, I., Lenior, M. and Cardon, G. (2013). 'Does a cycle training course improve cycling skills in children?', *Accident Analysis and Prevention*, **59**, 38–45.

Education Endowment Foundation and The Sutton Trust (2013). *Sutton Trust - EEF Teaching and Learning Toolkit* [online]. Available: http://educationendowmentfoundation.org.uk/uploads/toolkit/Teaching_and_Learning_Toolkit_(Spring_2013).pdf [4 December, 2014].

McLaughlin, K. and Glang, A. (2010). 'The effectiveness of a bicycle safety program for improving safety-related knowledge and behavior in young elementary students', *Journal of Pediatric Psychology*, **35**, 4, 343–353.

Turner, S., Roozenburg, A. and Smith, A. (2009). *Roundabout Crash Prediction Models* (Research Report No. 386). Wellington: NZ Transport Agency [online]. Available: http://www.nzta.govt.nz/resources/research/reports/386/docs/386.pdf [3 December, 2014].

# NFER provides evidence for excellence through its independence and insights, the breadth of its work, its connections, and a focus on outcomes.

- independent
- insights
- breadth
- connections
- outcomes